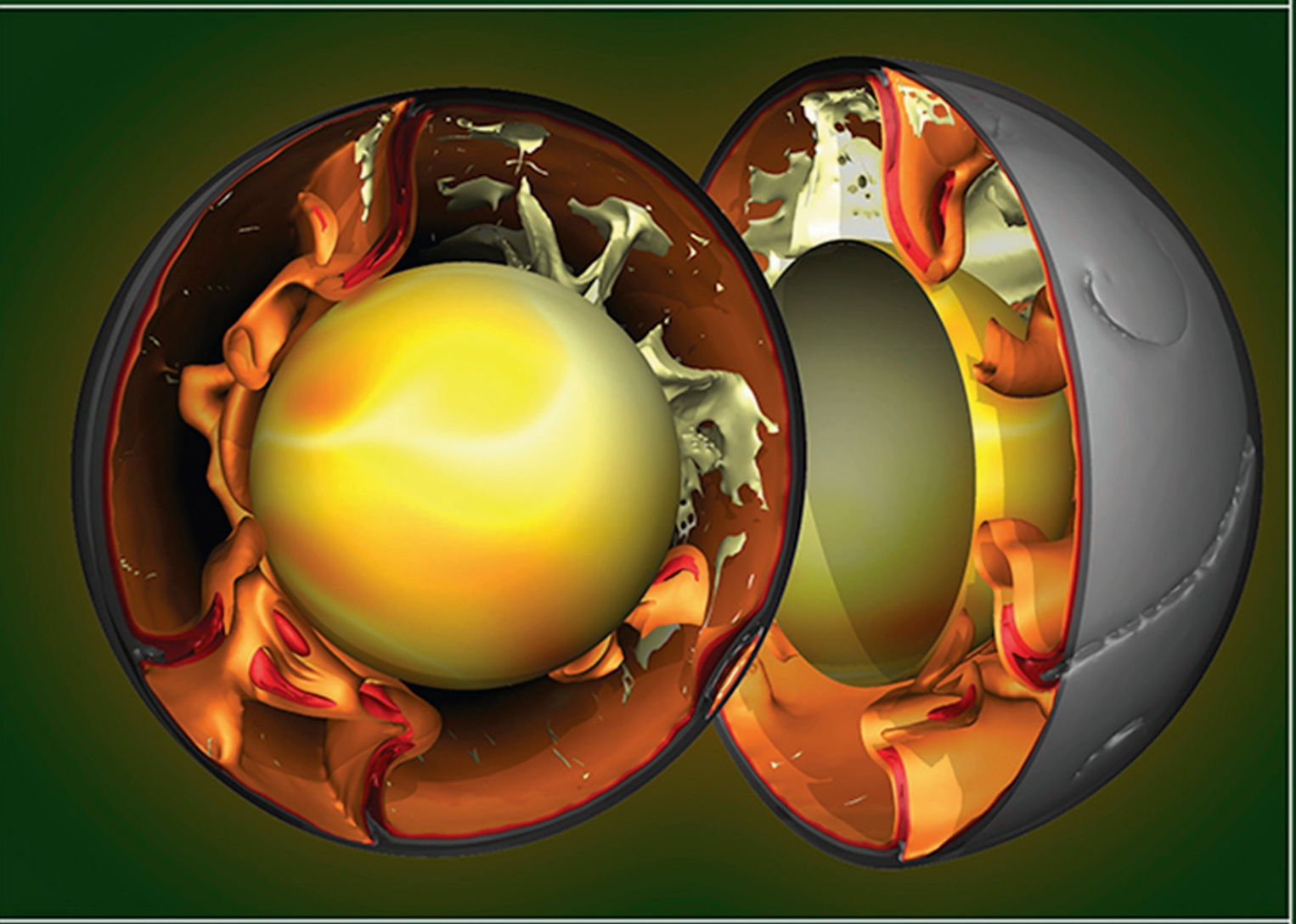
SECOND EDITION

TREATISE ON GEOPHYSICS

Volume 10

PHYSICS OF TERRESTRIAL PLANETS AND MOONS



VOLUME EDITOR: TILMAN SPOHN
EDITOR-IN-CHIEF: GERALD SCHUBERT

TREATISE ON GEOPHYSICS

SECOND EDITION

This page intentionally left blank

TREATISE ON GEOPHYSICS

SECOND EDITION

EDITOR-IN-CHIEF

Gerald Schubert

Distinguished Professor Emeritus, Department of Earth, Planetary and Space Sciences, University of California, Los Angeles

VOLUME 10

PHYSICS OF TERRESTRIAL PLANETS AND MOONS

VOLUME EDITOR

Professor Tilman Spohn

Institut für Planetenforschung, Deutsches Zentrum für Luft- und Raumfahrt (DLR)

ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington MA 01803, USA

Copyright © 2015 Elsevier B.V. unless otherwise stated. All rights reserved.

Cover image created for this work by Fabio Crameri, University College London.
The figure depicts the temperature field from a convection calculation using the code STAG-YY, written by Paul Tackley.

The following chapters are US Government works in the public domain and not subject to copyright:

1.11 Crust and Lithospheric Structure - Global Crustal Structure; 3.09 Earth Rotation Variations – Long Period; 4.07 The Role of Fault-Zone Drilling; 5.11 The Time-Averaged Field and Paleosecular Variation; 10.21 Exoplanetary Geophysics: An Emerging Discipline; 11.12 Tools and Techniques: Gravitational Method.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN (print): 978-0-444-53802-4

For information on all Elsevier publications
visit our website at http://store.elsevier.com/

Editorial & Production: Gemma Tomalin, Richard Berryman, Marise Willis, Michael Nicholls, Ellen Bishop
Cover design: Matthew Limbert

Printed and bound in the United Kingdom

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface *vii*

Editor-in-Chief *ix*

Contributors *xi*

Contents of all Volumes *xiii*

10.01 Physics of Terrestrial Planets and Moons: An Introduction and Overview 1
T Spohn

10.02 Interior Structure, Composition, and Mineralogy of the Terrestrial Planets 23
F Sohl and G Schubert

10.03 Planetary Seismology 65
P Lognonné and CL Johnson

10.04 Rotation of the Terrestrial Planets 121
T Van Hoolst

10.05 Gravity and Topography of the Terrestrial Planets 153
MA Wieczorek

10.06 Planetary Magnetism 195
JEP Connerney

10.07 Planetary Dynamos 239
FH Busse and RD Simitev

10.08 Dynamics and Thermal History of the Terrestrial Planets, the Moon, and Io 255
D Breuer and WB Moore

10.09 Planetary Tectonics and Volcanism: The Inner Solar System 307
TKP Gregg

10.10 Exogenic Dynamics, Cratering, and Surface Ages 327
SC Werner and BA Ivanov

10.11 Water on the Terrestrial Planets 367
J Helbert, E Hauber, and D Reiss

10.12 Solid Planet–Atmosphere Interactions 411
M Yu Zolotov

10.13 Planetary Atmospheres 429
DC Catling

10.14 Geology, Life, and Habitability 473
G Southam, F Westall, and T Spohn

10.15 Asteroids and Comets 487
YR Fernández, J-Y Li, ES Howell, and LM Woodney

10.16 Giant Planets 529
T Guillot and D Gautier

10.17 The Origin of the Natural Satellites 559
SJ Peale and RM Canup

10.18 Interiors and Evolution of Icy Satellites 605
H Hussmann, C Sotin, and JI Lunine

10.19 Introduction to 'Pluto, Charon, and the Kuiper Belt Objects': Pluto on the Eve of the *New Horizons* Encounter 637
WB McKinnon

10.20 Pluto, Charon, and the Kuiper Belt Objects 653
SA Stern, J Wm Parker, and CB Olkin

10.21 Exoplanetary Geophysics: An Emerging Discipline 673
G Laughlin and JJ Lissauer

10.22 Mission Analysis Issues for Planetary Exploration Missions 695
Y Langevin

10.23 Instrumentation for Planetary Exploration Missions 719
P Falkner and R Schulz

PREFACE

The *Treatise on Geophysics* is a comprehensive and in-depth study of the physics of the Earth and of terrestrial-like planets residing both within and outside our solar system. Its breadth and detail of coverage are beyond what any single geophysics text can provide. The first edition of the *Treatise on Geophysics* was published in 2007, nearly a decade ago. Of course, there is much progress in science in that length of time and the field of geophysics has grown rapidly and developed new lines of inquiry. Accordingly, we have brought the *Treatise on Geophysics* up-to-date with this second edition. The new edition will continue to provide students and professionals with fundamental and state-of-the-art discussion of all aspects of geophysics. In this new edition, the reader will find updates to all the chapters contained in the first edition and new chapters that discuss topics missing from the first edition. In a few instances, chapters from the first edition have simply been reprinted in order to avoid gaps in coverage. Chapters that emphasized fundamental physics often required little updating.

A highlight of the second edition is a new volume on Resources in the Near-Surface Earth. The new volume discusses the role of geophysics in the exploitation and conservation of natural resources and the assessment of degradation of natural systems by pollution. The near surface is a zone where humans and natural systems interact. Understanding the effects of human impacts is a challenge particularly for the long term.

There are four new chapters in the volume on Deep Earth Seismology. One of them describes progress in seismic instrumentation in the last 30 years in the era of broadband digital sensor and recording systems. Another one addresses a new approach for the determination of crust and uppermost mantle structure, based on ambient noise cross-correlations. This has revolutionized the field, since it allows the study of regional structure at high resolution in areas where there are no earthquakes. The third new chapter is devoted to recent advances in characterizing the lithosphere–asthenosphere boundary under both oceans and continents. Finally, the last new chapter discusses the morphology of subducted slabs in the Earth's transition zone.

The Mineral Physics volume has new chapters on dynamic (shock-wave) compression methods, and on the mineralogy of the upper mantle and the D″ layer focusing on post-perovskite, and extrasolar super-Earths.

New material in the Geodesy volume expands the temporal resolution and latency of geodetic measurements to high-rate, real-time measurements and increases the spatial resolution to meter size pixels with optical and radar imaging systems. Application areas are extended with new material on the uses of Earth gravity measurements.

The Earthquake Seismology volume has a new chapter on rock mechanics that is a comprehensive account of modern rock mechanics with emphasis on the variation of frictional properties of rocks during an earthquake cycle. Many chapters have considerable updates reflecting the lessons from significant recent earthquakes (e.g., the 2011 Tohoku-Oki earthquake and the 2013 Okhotsk deep earthquake), new observations on slow slips and triggered earthquakes, and developments of new methods (e.g., back projection method). Some chapters are updated (e.g., paleoseismology and strong ground motion) with substantial new material.

Chapters in the Geomagnetism volume have been updated to include recent advances in the acquisition of larger data sets and higher quality data and in the construction of more reliable and sophisticated models. Improvements in data have occurred in areas such as magnetic anomalies, paleointensities, and planetary magnetism. Improvements in models of the present day field include so-called comprehensive models and models that incorporate data assimilation.

The Crustal and Lithosphere Dynamics volume has new contributions on mountain building and landscape evolution.

The Mantle Dynamics volume contains new chapters on the generation of plate tectonics from mantle convection and the core–mantle boundary region.

The Core Dynamics volume features a comprehensive update of this rapidly changing subject, including new chapters on turbulence and waves in the fluid outer core.

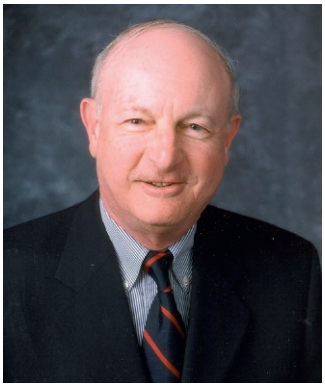
The volume on the Evolution of the Earth presents a new assessment on the role of volatiles in Earth evolution.

There are new chapters in the Physics of Terrestrial Planets and Moons volume on planetary atmospheres, planetary tectonics, extrasolar planets and asteroids and comets. The study of extrasolar planets has had explosive growth in the last decade and new spacecraft missions have significantly enhanced our knowledge of solar system planets and moons.

Finally, this is an opportunity for me to thank and acknowledge the editors of the individual volumes for their hard work, patience, dedication, and commitment in bringing this new edition of the Treatise to fruition. I am also indebted to Gemma Tomalin of Elsevier for her care and attention to the organization and production of the Treatise and to Deepa Arumugam for her careful coordination of the proofs from the copyeditor and typesetter. Of course, the authors of the individual chapters are the ones who are responsible for the excellence of the Treatise. We have all had some ups and downs in the process of producing the second edition of the Treatise but the final product is one that we can all view with pride.

Gerald Schubert
Editor-in-Chief

EDITOR-IN-CHIEF



Gerald Schubert is Distinguished Professor Emeritus in the Department of Earth, Planetary and Space Sciences at the University of California, Los Angeles. His research interests encompass the physics of the interiors and atmospheres of the Earth, the Moon, and the other moons and planets of the solar system. He is co-author with Donald Turcotte of *Geodynamics* (ed. 3, Cambridge University Press, 2014) and with Turcotte and Peter Olson of *Mantle Convection in the Earth and Planets* (Cambridge University Press, 2001). He is the author of over 540 research papers. He has participated in a number of NASA's planetary missions, including Apollo, Pioneer Venus, Magellan, and Galileo, and has been editor and editorial board member of many journals, including *Icarus*, *Journal of Geophysical Research*, *Geophysical Research Letters*, and *Annual Reviews of Earth and Planetary Sciences*. Professor Schubert is a Fellow of the American Geophysical Union and a recipient of the Union's James B. MacElwane medal and the Harry H. Hess medal. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

This page intentionally left blank

CONTRIBUTORS

D Breuer
DLR, Institut für Planetenforschung, Berlin, Germany

FH Busse
Physikalisches Institut der Universität Bayreuth, Bayreuth, Germany

RM Canup
Southwest Research Institute, Boulder, CO, USA

DC Catling
University of Washington, Seattle, WA, USA

JEP Connerney
NASA Goddard Space Flight Center, Greenbelt, MD, USA

P Falkner
European Space Agency, Noordwijk, The Netherlands

YR Fernández
University of Central Florida, Orlando, FL, USA

D Gautier
LESIA, Observatoire de Paris, Meudon, France

TKP Gregg
University at Buffalo, Buffalo, NY, USA

T Guillot
Observatoire de la Côte d'Azur, CNRS, Université de Nice-Sophia Antipolis, Nice, France

E Hauber
Institute for Planetary Research, DLR, Berlin, Germany

J Helbert
Institute for Planetary Research, DLR, Berlin, Germany

ES Howell
Arecibo Observatory/USRA, Arecibo, PR, USA

H Hussmann
DLR Institute of Planetary Research, Berlin, Germany

BA Ivanov
Institute for Dynamics of Geospheres, Moscow, Russia

CL Johnson
University of British Columbia, Vancouver, BC, Canada; Planetary Science Institute, Tucson, AZ, USA

Y Langevin
CNRS/Université Paris Sud, Orsay, France

G Laughlin
University of California at Santa Cruz, Santa Cruz, CA, USA

J-Y Li
Planetary Science Institute, Tucson, AZ, USA

JJ Lissauer
NASA Ames Research Center, Moffett Field, CA, USA

P Lognonné
Université Paris Diderot -Sorbonne Paris Cité, Institut de Physique du Globe de Paris, Paris, France

JI Lunine
Cornell University, Ithaca, NY, USA

WB McKinnon
Washington University in St Louis, Saint Louis, MO, USA

WB Moore
Hampton University, Hampton, VA, USA

CB Olkin
Southwest Research Institute, Boulder, CO, USA

J Wm Parker
Southwest Research Institute, Boulder, CO, USA

SJ Peale
University of California, Santa Barbara, CA, USA

D Reiss
Westfälische Wilhelms-Universität Münster, Münster, Germany

G Schubert
University of California, Los Angeles, CA, USA

R Schulz
European Space Agency, Noordwijk, The Netherlands

RD Simitev
University of Glasgow, Glasgow, UK

F Sohl
German Aerospace Center (DLR), Berlin, Germany

C Sotin
JPL, California Institute of Technology, Pasadena, CA, USA

G Southam
The University of Queensland, Brisbane, QLD, Australia

T Spohn
DLR Institut für Planetenforschung; German Aerospace Center (DLR) Berlin, Germany

SA Stern
Southwest Research Institute, Boulder, CO, USA

T Van Hoolst
Royal Observatory of Belgium, Brussels, Belgium; KU Leuven, Leuven, Belgium

SC Werner
University of Oslo, Oslo, Norway

F Westall
Centre de Biophysique Moléculaire, Orléans, France

MA Wieczorek
University Paris Diderot, Paris Cedex 13, France

LM Woodney
California State University – San Bernardino, San Bernardino, CA, USA

M Yu Zolotov
Arizona State University, Tempe, AZ, USA

10.01 Physics of Terrestrial Planets and Moons: An Introduction and Overview

T Spohn, DLR Institut für Planetenforschung, Berlin, Germany

© 2015 Elsevier B.V. All rights reserved.

10.01.1	Introduction	1
10.01.2	Our Planetary System	1
10.01.3	Planetary Missions	4
10.01.4	Planet and Satellite Orbits and Rotation States	8
10.01.5	Composition and Interior Structure of Planets	9
10.01.6	Surfaces and Atmospheres	12
10.01.7	Energy Balance and Evolution	15
10.01.8	Magnetic Fields and Field Generation	17
10.01.9	Origin of the Solar System	19
10.01.10	Concluding Remarks	20
References		21

10.01.1 Introduction

Humanity has always been fascinated with the wandering stars in the sky, the planets. Ancient astrologists have observed and used the paths of the planets in the sky to time the seasons and to predict the future. Observations of the planets helped J. Kepler to formulate his laws of planetary motion and revolutionize the perception of the world. With the advent of the space age, the planets have been transferred from bright spots in the sky to worlds of their own right that can be explored, in part by using the in situ and remote sensing tools of the geosciences. The terrestrial planets are of particular interest to the geoscientists because comparison with our own planet allows a better understanding of our home, the Earth. Venus offers an example of a runaway greenhouse that has resulted in what we would call a hellish place. With temperatures of around 450 °C and a corrosive atmosphere that is also optically nontransparent, Venus poses enormous difficulties to spacecraft exploration. Mars is a much friendlier planet to explore but a planet where greenhouse effects and atmospheric loss processes have resulted in a cold and dusty desert. But aside from considerations of the usefulness of space exploration in terms of understanding the Earth, the interested mind can visit astounding and puzzling places. There is the dynamic atmosphere of Jupiter with a giant thunderstorm that has been raging for centuries. There is Saturn with its majestic rings, and there are Uranus and Neptune with complicated magnetic fields. These giant planets have moons that are astounding. There is the volcanic satellite of Jupiter, Io, that surpasses the Earth and any other terrestrial planet in volcanic activity. This activity is powered by tides that twist the planet such that its interior partially melts. A much smaller moon of Saturn, Enceladus, also has geysers that could be powered by tidal heating although the case is less clear. Its volcanic activity releases water vapor not lava, however. There is another moon of Saturn, Titan, that hides its surface underneath a layer of photochemical smog in a thick nitrogen atmosphere, and there are moons of similar sizes that lack any comparable atmosphere. Miranda, a satellite of Uranus, appears as if it has been ripped apart and reassembled. And Triton, the major satellite of Neptune, has geysers of nitrogen powered by solar irradiation. Magnetic field data suggest that icy moons orbiting the giant planets may have oceans underneath thick ice covers. These oceans can, at least in principle, harbor or have harbored life. Moreover, there are asteroids with moons and comets that may still harbor the clues to how the solar system and life on Earth formed.

This volume of the Treatise on Geophysics discusses fundamental aspects of the science of the planets. It is focused on geophysical properties of the Earth-like planets and moons, those bodies that consist largely of rock, iron, and water, and the processes occurring in their interiors and on their surfaces. But it goes further by discussing the giant planets and asteroids and comets as well and other extrasolar planetary systems. The better part of the volume is dedicated to the formation, interior structure, and evolution of the terrestrial planets and to their physical properties such as gravity and magnetic fields, rotation, the physics of the atmosphere, and surface–atmosphere interactions. What is the planetological context of life and are there feedbacks between life and planetary evolution? An attempt at answering these questions is the subject of a dedicated chapter. In the second part, we turn to the outer fringes of the solar system beyond the asteroid belt, to the giant planets and their satellites, to Pluto and the Kuiper Belt, and, finally, to extrasolar planets. Because exploration by spacecraft is so fundamental to the geophysics of the planets, the volume has two chapters that discuss the orbital dynamics – the routes to the planets – and the instruments typically flown on planetary missions. This edition of the volume has been widened in scope and coverage in comparison with the previous editions with dedicated chapters on asteroids and comets, on planetary tectonics and volcanism, on planetary atmospheres, and on extrasolar planetary systems.

10.01.2 Our Planetary System

The solar system contains a myriad of bodies ranging in size from the Sun to miniscule dust particles. The *Encyclopedia of Planetary Sciences* (Shirley and Fairbridge, 1997; on the web as part of the *Earth Sciences Series* at www.springerreference.com), *Encyclopedia*

of Astrobiology (Gargaud et al., 2011), and *Encyclopedia of the Solar System* (Spohn et al., 2014); Vol. VI-4-B of the Landolt–Börnstein Series (Trümper, 2009); the Planetary Scientist's Companion (Lodders and Fegley, 1998); and textbooks such as de Pater and Lissauer (2010) are useful sources of information about the solar system. Valuable collections of planetary data can be found on http://science.jpl.nasa.gov/PlanetaryScience, on http://ssd.jpl.nasa.gov, and on http://nssdc.gsfc.nasa.gov. A collection of images of planetary surfaces can be found in the NASA Photojournal (http://photojournal.jpl.nasa.gov).

The eight planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune are shown in **Figure 1** together with Pluto, the most prominent member of a new class of objects, the dwarf planets. The International Astronomical Union (IAU) introduced the class of dwarf planets in 2006. These are intermediate in size between the terrestrial planets and the 'small solar system bodies' such as cometary nuclei and most asteroids. The IAU has so far identified the following five celestial bodies as dwarf planets – Pluto, the asteroid Ceres, and the trans-Neptunian objects (TNO) Eris, Haumea, and Makemake – but there should be many more still to be recognized dwarf planets in the solar system (e.g., http://en.wikipedia.org/wiki/List_of_possible_dwarf_planets). A planet of our solar system according to the IAU resolution 5 (http://www.iau.org/news/pressreleases/detail/iau0603/#1) is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighborhood around its orbit. A dwarf planet according to the resolution satisfies (a) and (b) but has *not* cleared its neighborhood and (d) is not a satellite. All other objects, except satellites – so the resolution continues – shall be referred to collectively as 'small solar system bodies.' See also Basri and Brown (2006) for a discussion of the term 'planet.'

The Sun, a middle-aged main sequence star, contains 98.8% of the mass of the solar system but only 0.5% of its angular momentum. The next smaller body, Jupiter, still 300 times more massive than the Earth (see **Table 1**), contains more than 60% of the mass of the rest. Jupiter is the biggest of the *giant planets*, a group of gaseous planets that constitute a major subgroup of the solar system and other known planetary systems (see **Chapters 10.16** and **10.21**). Among the giant planets are, in addition to Jupiter, Saturn, Uranus, and Neptune. The latter two are sometimes called the subgiants or the ice giants because they are notably smaller than Saturn and Jupiter and because they mostly consist of water, methane, and ammonia, components often collectively called the planetary ices (see below). The Earth, the biggest member of the other major subgroup, the *terrestrial planets*, is the only planet on which we know to date that life has originated and evolved. Among the members of this group are Mercury, the innermost planet; Venus, the Earth's twin with respect to size and mass; and Mars. The latter planet has the best chance of having – or having once had – some primitive forms of life, which makes it a prime target of present-day space missions. The terrestrial planets together have about 0.005% of the mass of the solar system. **Table 1** collects some data of general interest on the planets, dwarf planets, and some major moons.

Comparative planetology is the science of studying the planets by comparing and finding general properties and common lines of evolution as well as features that are specific. It uses the methods of the natural sciences and is strongly interdisciplinary. Comparative planetology is not restricted to the planets *sensu stricto* but also considers the major moons of the planets such as the Earth's Moon, the major satellites of Jupiter (Io, Europa, Ganymede, and Callisto; **Figure 2**), the major Saturnian satellite Titan, and, finally, Triton, Neptune's major satellite. Other bodies of interest include the yet largely unexplored members of the asteroid belt, the asteroids, the comets, and the Kuiper Belt

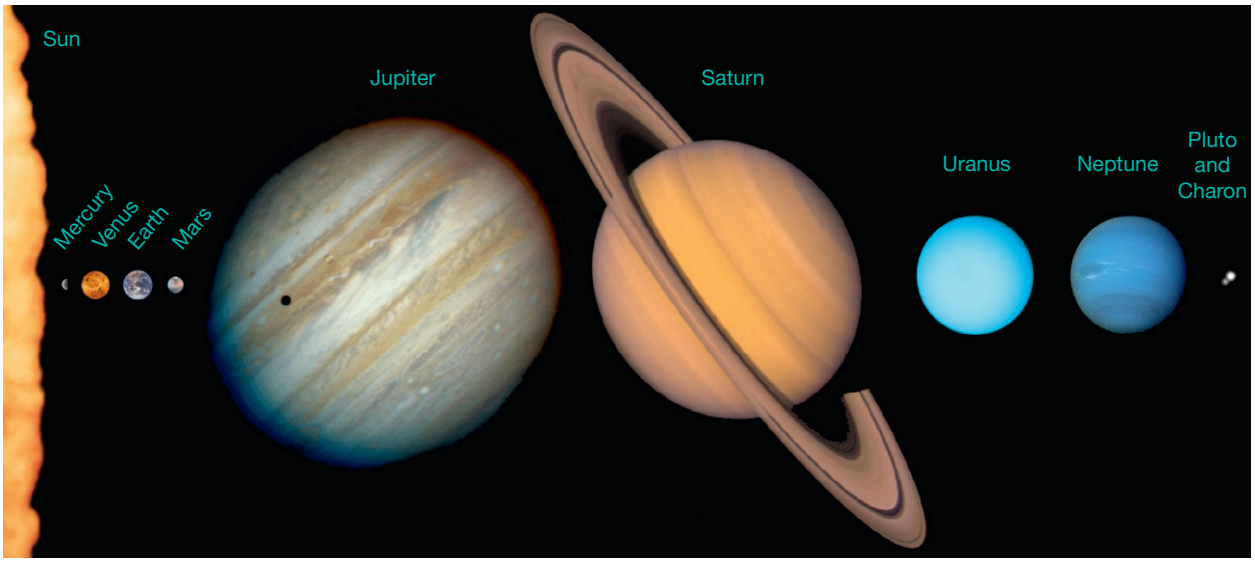


Figure 1 The eight planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune and the dwarf planet Pluto are shown in this compilation of NASA images with their correct relative sizes and ordered according to their distance from the Sun. Mercury is barely visible at left close to the arc of the surface of the Sun, and Pluto is barely visible at the outer right. The dark spot on Jupiter is the shadow of Io, one of its major satellites. © C. J. Hamilton.

Table 1 Some data of general interest on the planets, dwarf planets, and some major moons

(a) Properties of the planets

	Terrestrial or Earth-like planets				*Giant planets*			
	Mercury	*Venus*	*Earth*	*Mars*	*Jupiter*	*Saturn*	*Uranus*	*Neptune*
Radius (km)	2438.0	6052.0	6371.0	3390.0	71492.0	60268.0	24973.0	24764.0
Mass (10^{24} kg)	0.3301	4.869	5.974	0.6419	1899.0	568.46	86.83	102.4
Density (10^3 kg m^{-3})	5.430	5.243	5.515	3.934	1.326	0. 6873	1.318	1.638
Uncompressed density (kg m^{-3})	5.3	4.0	4.05	3.75	0.1	0.1	0.3	0.3
Rotational period (d[a])	58.65	243.0[b]	0.9973	1.026	0.4135	0.4440	0.7183[b]	0.6713
Inclination of rotation axis (°)	0.03	177.3	23.44	25.19	3.13	26.73	97.8	28.32
Orbital distance (AU[c])	0.3871	0.7233	1.000	1.524	5.203	9.572	19.19	30.07
Orbital period (a[d])	0.2410	0.6156	1.001	1.882	11.87	29.39	84.16	165.0
Magnetic moment (10^{-4} T $\times$ radius^3)	3×10^{-3}	$<3\times10^{-4}$	0.61	$<6\times10^{-4}$	4.3	0.21	0.23	0.133
Effective surface temperature (K)	445	325	277	225	123	90	63	50
Specific heat flow or luminosity (pW kg^{-1})	?	?	7	?	176	152	4	67
Known satellites	0	0	1	2	65	62	27	13

(b) Properties of dwarf planets (ordered by their distance from the Sun)

	Ceres	*Pluto*	*Haumea*	*Makemake*	*Eris*
Radius (km)	476.2	1151.0	620	715	1200 ± 50
Mass (10^{24} kg)	9.47×10^{-4}	0.01309	?	?	?
Density (10^3 kg m^{-3})	2.09	2.050	?	?	?
Rotational period (d[a])	0.378	-6.387[b]	0.163	?	?
Orbital distance (semi major axis) (AU[c])	2.765	39.48	43.00	45.34	67.73
Orbital period (a[d])	4.6033	248.09	282,12	305,55	561.74
Orbital eccentricity	0.07913	0.249	0.19813	0.16384	0.4336
Orbital inclination (°)	10.59	17.14	28.22	29.00	46.87
Effective surface temperature (K)	167	44	34	36	$\sim$30
Known satellites	0	5 (Charon, Nix, Hydra, Kerberos, Styx)	2 (Namaka, Hi'iaka)	0	1 (Dysnomia)

(c) Properties of major satellites

	Moon	*Io*	*Europa*	*Ganymede*	*Callisto*	*Titan*	*Triton*
Primary	Earth	Jupiter				Saturn	Neptune
Radius (km)	1737.4	1821.5	1562.1	2632.3	2409.3	2575.5	1352.6
Mass (10^{20} kg)	734.9	893.2	480.0	1481.9	1075.9	1346.0	214.0
Density (10^3 kg m^{-3})	3.344	3.53	3.02	1.94	1.85	1.881	2.061
Specific heat flow (pW kg^{-1})	8	890	150(?)	?	?	?	?
Orbital period (d)	27.32	1.769	3.551	7.155	16.69	15.95	5.877[b]

For the sake of space, we give four significant digits even for numbers that are more accurately known.
The uncompressed density is model dependent and cannot be given with more than 2–3 significant digits.
[a]A day (d) is equivalent to 24 h.
[b]The motion (rotation, revolution) is retrograde.
[c]AU is one astronomical unit or 149.6 million kilometers.
[d]A year (a) is equivalent to 365 days (d).
Source: The data have been taken from the compilation of Lodders and Fegley, The Planetary Companion, Oxford University Press, 1998, from Spohn, Planetologie, in Bergmann-Schäfer, Lehrbuch der Experimentalphysik, Vol. 7, (ed. W. Raith), W. De Gruyter, Berlin, p. 427–525 and from http://nssdc.gsfc.nasa.gov and http://ssd.nasa.gov.

objects. The asteroid belt is found between Mars and Jupiter, and its members are generally accepted to be the parent bodies of most meteorites (stones from space found on the Earth's surface; see Chapter 10.15). The Kuiper Belt stretches between 30 AU (roughly the orbit of Neptune) and at least 50 AU and contains the dwarf planets Pluto, Haumea, Makemake, and Eris (Chapters 10.19 and 10.20). The asteroids are believed to be the source of most meteorites and contain rich information about the formation of the solar system and its evolution and even the evolution of matter before the formation of the solar system. The Kuiper Belt objects are largely unexplored. Pluto, the most prominent member, and the Kuiper Belt are the target of the New Horizons mission launched in 2006 and scheduled to arrive at Pluto in 2015 (see Chapters 10.19 and 10.20). The emerging field

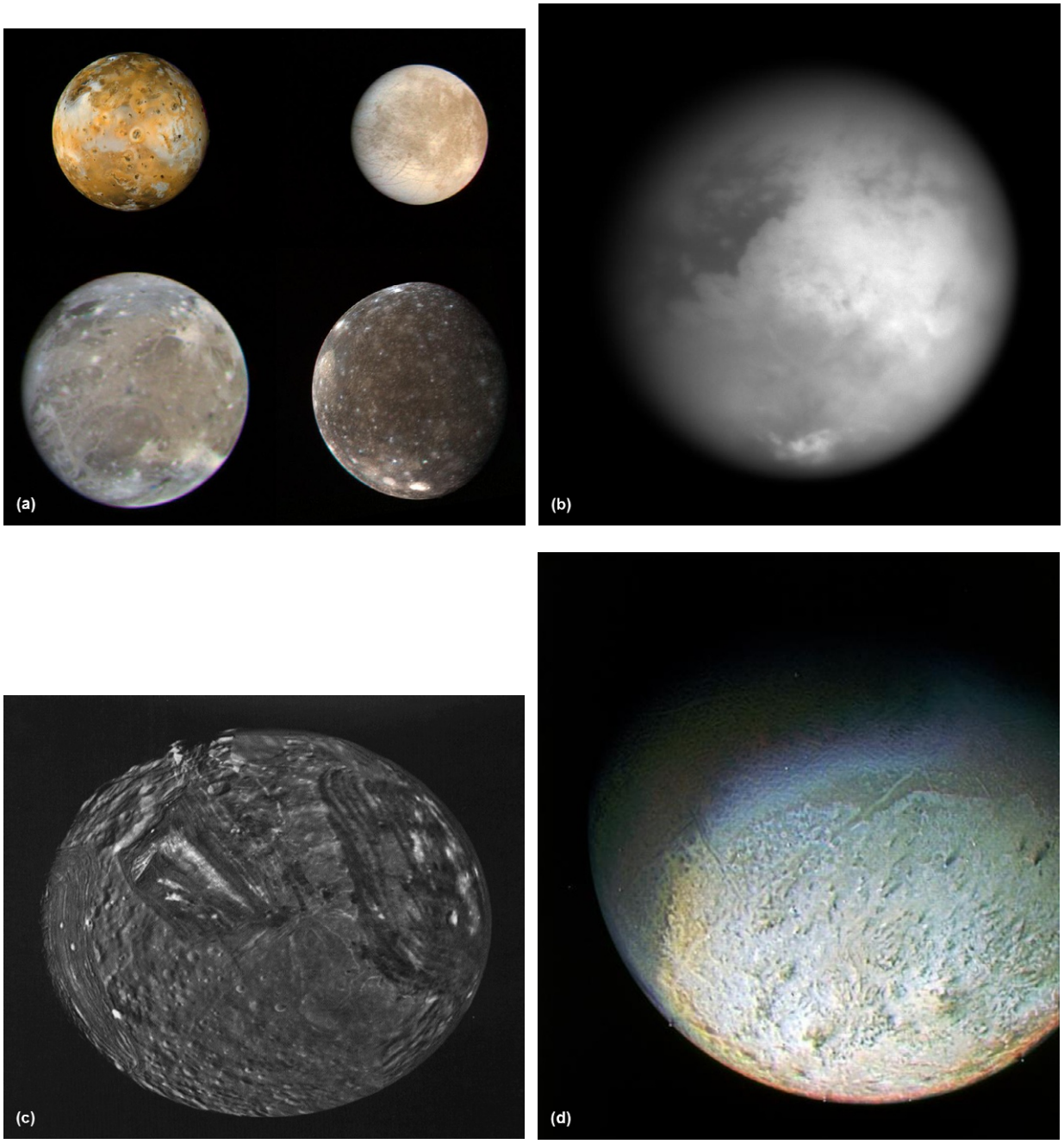


Figure 2 The four Galilean satellites of Jupiter Io, Europa, Ganymede, and Callisto (a), the Saturnian satellite Titan (b), the Uranian satellite Miranda (c), and the Neptunian satellite Triton (d) are shown in this compilation of spacecraft images from top left to bottom right. NASA/JPL.

of extrasolar planetology will further widen the scope of comparative planetology.

10.01.3 Planetary Missions

The results of planetary sciences have been made possible at their present levels only through the exploration of the solar system with space missions (see, e.g., Burke, 2014 and **Chapters 10.22** and **10.23**. See also a list of missions to the planets in Spohn et al. (2014)). All planets and major satellites of the solar system have been visited at least by a flyby of a robot spacecraft. This is not true of the dwarf planets, but missions are on their ways to Pluto (New Horizons) and Ceres (Dawn). Dawn has orbited Vesta between July 2011 and September 2012 (Russel and Raymond, 2011 and **Chapter 10.15**) and is planned to go into orbit around Ceres in February 2015. New Horizons is planned to flyby Pluto and its satellites with closest approach in June 2015 (**Chapters 10.19** and **10.20**). Some, such as Venus, the Moon, Mars, Jupiter, and Saturn, have been

explored by orbiters and by landers. The Moon has even been visited by astronauts. It can be said without any doubt that space exploration has turned dim disks in the sky observable with telescopes into worlds of their own rights waiting for further, perhaps eventually, human exploration.

Space exploration began with missions to the Moon in 1959, characterized by the race to the Moon between the United States and the then Soviet Union. This race culminated in the landing of men on the Moon. For planetary sciences, the significance of the Apollo program lies mainly with the first sample return, the first seismic exploration of a planetary interior other than the Earth's, the first heat flow measurement, and the first geologic field trip on another planet. Although Apollo has not finally solved the puzzle of the origin of the Moon – some may argue that it can never be solved once and for all – it has provided a wealth of data and has made the Moon one of the best-known planetary bodies. Scientists continue to successfully explore the Apollo data as the recent discovery of water in lunar rock (Saal et al., 2008) and the seismic detection of the core (Weber et al., 2011) have demonstrated. Missions to the Moon have been accomplished by Japan, China, India, and ESA (Kaguya, Chang'e, Chandrayaan, and Smart-1). A recent highlight of geophysical exploration of the Moon has been the GRAIL (Gravity Recovery and Interior Laboratory) mission of NASA with two spacecraft doing gravity gradiometry (**Chapter 10.05**).

Highlights of Mars exploration include the Viking missions, the first landings on Mars; the Pathfinder mission, the first vehicle introducing mobility; Mars Express, the first European planetary mission *sensu stricto*; the Mars Exploration Rovers, MER for short; and the Mars Science Laboratory on the Curiosity rover. From a geophysics point of view, Mars Global Surveyor, which returned highly accurate maps of topography and gravity (Smith et al., 1999) and detected the remnant magnetization of old crustal units (Acuña et al., 1999), has been a particularly remarkable mission. The MER (**Figure 3**) have been doing the first geologic field trips on Mars for the past 10 years and detected on their now tens of kilometers long journey minerals that formed under humid conditions (Squyres and Knoll, 2005 and papers therein). The Mars Express mission detected methane on Mars, numerous morphological units that required water to form (**Figure 4**, **Chapter 10.11**) and showed that volcanic activity continued albeit at a low rate until the recent past (Neukum et al., 2004). Mars Science Laboratory/Curiosity, more massive than any previous planetary exploration rover and equipped with a most sophisticated payload, is exploring the Gale crater and has found evidence for a past environment well suited to supporting microbial life. The first *in situ* geophysical observatory InSight equipped with a very broad band seismometer (compare **Chapter 10.03**), a heat flow probe, and a magnetometer is scheduled for launch in 2016. The lander will use its communication hardware for radio science tracking of, for example, the rotation axis of the planet. A sample return mission to Mars continues to be high on the agenda of the planetary science community.

An important driver for Mars research is its potential habitability and the possible detection of life, extinct or extant. The Viking landers in the 1970s already attempted to find signatures of bioactivity but for most failed to do so. Beagle 2 onboard the European Mars Express Mission was built to search for bioactivity and would have been the first European lander on another planet had it not failed. The next attempt to directly detect biosignatures on Mars will be the ESA ExoMars rover rescheduled for launch in 2018.

Venus has been the target of very successful Soviet missions that included eight Landers (Venera 7–14). These landers survived in the highly corrosive atmosphere of Venus only for a

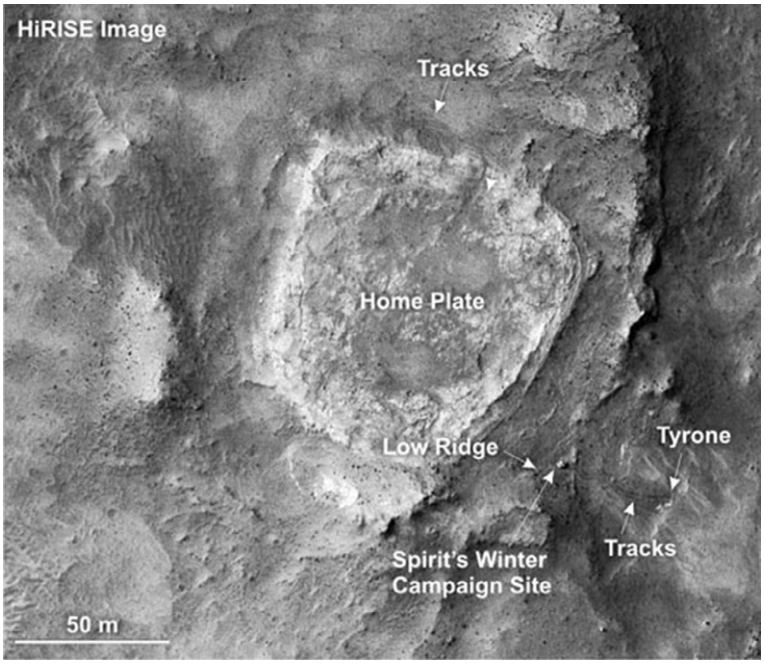


Figure 3 Traces of the MER rovers imaged by the ultimate-resolution camera onboard the Mars Reconnaissance Orbiter. NASA/JPL.

Figure 4 This image has been interpreted to show a frozen over lake on Mars comparable in size to the North Sea on Earth. © ESA/DLR/FU Berlin (G. Neukum).

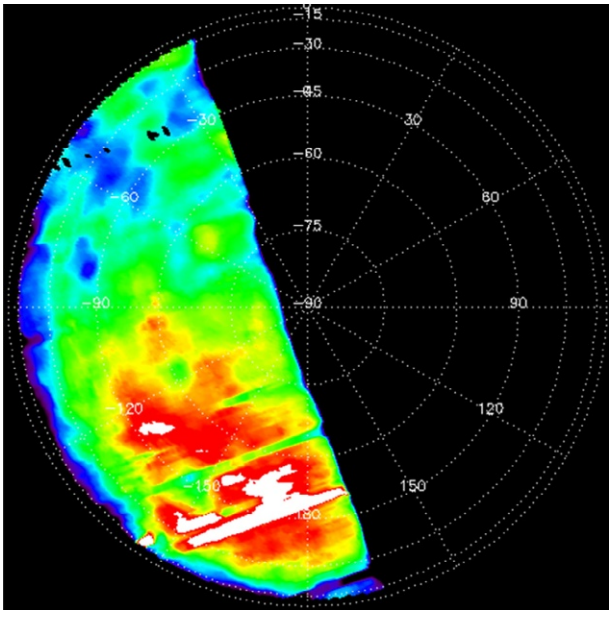


Figure 5 Map of half of the southern hemisphere of Venus in the infrared taken with the VIRTIS infrared spectrometer on Venus Express. © ESA.

few hours but transmitted color photos from the surface that shape our image of the Venusian surface. Further highlights of Venus exploration have been the Pioneer Venus, Magellan, and Venus Express missions, which provided extensive data on the atmosphere of Venus and the first systematic mapping of the surface with radar and in the infrared (Figure 5). Since the atmosphere of Venus is optically thick, cameras are of little use at this planet for surface exploration but are used to study atmosphere circulation patterns.

Mercury, the innermost planet, was visited by two flybys of the American Mariner 10 mission targeted for Mars in the 1970s. But since March 2011, the NASA MESSENGER mission has been orbiting the planet and has provided a wealth of data. The next mission to the planet will be the ambitious European BepiColombo mission scheduled for launch in 2016 with orbit insertion in 2024 and with a planetary remote sensing orbiter and a second orbiter – provided by the Japanese space agency JAXA – to explore the Hermean magnetosphere.

The outer solar system was first explored by the Pioneers that flew by Jupiter (Pioneers 10 and 11) and Saturn (Pioneer 11) and subsequently by the very successful Voyager I and II missions. The latter flew by all four major planets of the outer solar system. In addition to exploring the planets, these spacecraft discovered rings at Jupiter and Uranus and explored the satellite systems and found Io to be the most volcanically active body in the solar system. The Voyagers for the first time allowed a detailed view of the rich phenomenology of icy satellite surfaces. After the Pioneer and Voyager flybys, orbiters were the next logical step in outer solar system exploration. The first mission in this class was Galileo that orbited Jupiter and plunged into Jupiter's crushing atmosphere in September 2003. The spacecraft was deliberately destroyed to protect the Jovian system, in particular Europa, from being polluted. Galileo discovered the magnetic field of Ganymede and evidence for oceans underneath the ice crusts of the Galilean moons Europa, Ganymede, and Callisto. On its route to Jupiter, Galileo discovered for the first time a moon (Dactyl) that is orbiting an asteroid (Ida) and observed comet Shoemaker–Levy crashing into Jupiter. Galileo imaged the Jovian satellites at much improved resolution and discovered several new ones. Comparison with Voyager images of volcanic features on Io showed significant modifications that had occurred in the roughly 20 years between the two missions. An instrumented descent probe released from the Galileo orbiter entered the Jovian atmosphere in December 1995 and provided the first *in situ* measurements of the state and the chemistry of a giant planet atmosphere shroud. The results of the Galileo mission are discussed in Bagenal et al. (2004). The NASA Juno mission scheduled to go into a polar orbit in July 2016 to study the planet's gravity and magnetic fields, its atmosphere dynamics, and its composition will revisit Jupiter. The Jovian satellites will be the prime target of ESA's JUICE (JUpiter ICy moons Explorer) mission in the 2020s. Among the most spectacular planned missions to the Jovian system is a lander mission for Europa either as a Russian contribution to the JUICE mission or as a NASA mission. The mission plans are driven by the perception that an ocean on Europa may be covered by a thin ice shell that can be drilled and by the possibility that life developed in this ocean. It is possible that water from this ocean may have surfaced through cracks in the ice shell and may carry chemical signals of a biosphere.

At the time of this writing, Cassini has been orbiting Saturn for a decade. Cassini has even topped the success of Galileo as an outer solar system exploration mission through its more advanced payload capabilities and longer mission duration. The Huygens probe supplied by the European Space Agency (ESA) descended through the atmosphere of Saturn's biggest moon Titan and successfully landed on the surface on 14 January 2005 to transmit the first images (color) from Titan's surface (Figure 6). Titan's atmosphere is optically thick like that of Venus. Huygens landed on a solid surface strewn with ice boulders. On its descent, it imaged rivers in which

Figure 6 First color image from Titan's surface. This image was processed from images returned on 14 January 2005, by ESA's Huygens probe during its successful descent to land on Titan. The rocklike objects just below the middle of the image are about 10 cm across. The surface consists of a mixture of water and hydrocarbon ice. There is also evidence of erosion at the base of these objects, indicating possible fluvial activity. © ESA/NASA/JPL/ University of Arizona.

hydrocarbons are flowing or have flowed in the past and a possible shoreline (Lebreton et al., 2005). Near-infrared data from the Cassini mission have been interpreted as showing cryovolcanoes on Titan (e.g., Sotin et al., 2005). Another highlight of the mission is the detection of active geysers on the small satellite Enceladus (Porco et al., 2006, **Figure 7**), the diameter of which is only 500 km, and a rift structure circling Iapetus along its equator (Porco et al., 2005). Cassini also returned gravity data that allowed models of the interiors of some of the smaller satellites (see **Chapter 10.18**). Most recently, Cassini gravity data taken at Enceladus have been interpreted to support the existence of the long speculated (partial) ocean that could feed the geysers (Iess et al., 2014). This ocean would not be global but extend underneath most of the southerly hemisphere and has been speculated to be another potential habitat. Cassini will continue to explore the Saturn system through the Cassini Solstice Mission until September 2017 and then plunge into Saturn. Mission ideas to follow in the footsteps of Cassini include dedicated Titan and Enceladus Explorers. Just as for the Europa missions, these

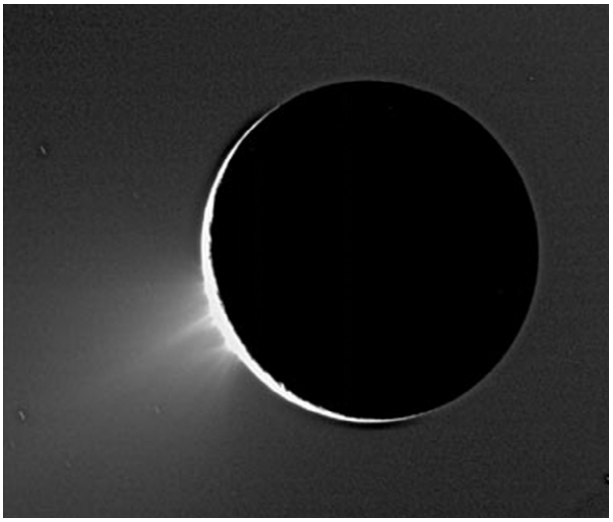


Figure 7 Geysers emanating from the southern hemisphere of Enceladus. NASA/JPL.

proposal are at least in part driven by the interest in possible extraterrestrial life.

Asteroids and comets are believed to be remnants of accretion and to consist of largely primitive matter (see **Chapter 10.15**). Although this perception is questionable to some extent – comet nuclei that have repeatedly flown by the Sun may have experienced thermal metamorphism; some asteroids show signs of endogenic activity and may be differentiated – asteroids and comets are minor solar system bodies of great interest. The Dawn mission has explored the second largest asteroid Vesta and is on its way to the dwarf planet Ceres. Vesta has been shown to be a differentiated body, while the latter could be undifferentiated and mostly primitive. Comet nucleus Churyumov–Gerasimenko is the target of the ESA Rosetta mission launched in 2004. After a 10-year cruise, the Rosetta orbiter and the lander Philae have ended hibernation and will begin to orbit the nucleus in late summer of 2014, while the lander Philae will dock onto the nucleus in the fall and explore it *in situ*. (We are speaking of docking rather than landing because of the small gravity of the nucleus.) Both the orbiter and Philae will take data as the comet approaches the Sun.

Planetary exploration has taken and is still following a stepped approach by which flybys are followed by orbiter missions. These are followed by lander missions including networks of landing stations and rovers. Sample return to the Earth is then foreseen, followed by human exploration. The latter step is often met with skepticism because of its prohibitive cost and is still facing huge problems of technical and medical nature and of safety for the astronauts.

The payloads of orbiters and flyby spacecraft are specifically designed for their science goals (see **Chapter 10.23**). It is still possible, however, to characterize generic payload elements that are typically found on these spacecraft. Almost every mission has at least one camera, often two for wide angle and for high resolution. Early missions had panchromatic cameras, but recent missions have color capabilities. Stereo is a useful feature that can increasingly be found on missions like Mars Express. Stereo allows 3-D imaging and enables the

development of digital terrain models. Digital terrain modeling is helped by laser altimetry as the successful Mars Observer Laser Altimeter (MOLA) on Mars Global Surveyor has so convincingly demonstrated. If the surface is hidden underneath an optically thick atmosphere, radar can be used to map the surface as has been successfully done at Venus by the Magellan mission. Radar is further used to map the surface roughness and to penetrate into the first centimeters of the near-surface layers, for instance, to look for ground ice. Other microwave sounders have been proposed that use different wavelengths and have differing depths of penetration.

While cameras, laser altimeters (**Figure 8**), and radio science instrumentation are indispensable instruments for planetary geology, geophysics, and geodesy, spectrometers are the tools of cosmochemistry and mineralogy. These instruments typically analyze particles (photons, neutrals, and ions) emitted from surfaces and atmospheres. Visible, near-infrared, and thermal infrared (0.5–15 μm) spectrometers are used to characterize the mineralogy of surface rock (silicates, carbonates, sulfates, etc.) and to search for water ice. Mapping infrared spectrometers such as Themis on Mars Odyssey and Omega on Mars Express allow a mapping of minerals on the surface. The Planetary Fourier Spectrometer on Mars Express works in a similar but broader spectral range (1.2–45 μm) but has a much higher spectral resolution at the expense of spatial resolution. Ultraviolet spectrometers measure neutral metals such as Al, Na, and S, ozone, and OH radicals. Gamma-ray spectrometers can measure radio nuclides (K, Th, and U), major rock-forming elements (O, Mg, Al, Si, Ca, Ti, and Fe), hydrogen (from water), and carbon, depending on the energy resolution and the elemental concentration in the surface. x-Ray spectrometers are suitable to detect the major elements Mg, Al, Si, Ca, and Fe.

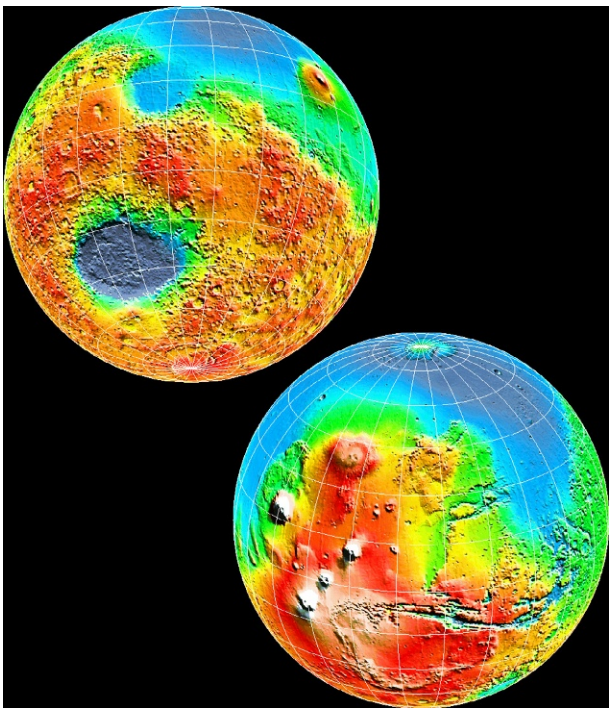


Figure 8 Global map of Mars taken by the Mars Observer Laser Altimeter (MOLA) onboard the Mars Global Surveyor mission. NASA/JPL.

The particles and electromagnetic fields surrounding planets and other solar system objects are measured with magnetometers and plasma detectors. Dust analyzers, finally, collect dust particles and analyze the direction and strength of dust flux in space.

Landing missions carry cameras and have been equipped with spectrometers and other sensors to measure composition and mineralogy. Mobility is an important issue since exploration of just one landing side may lead to biased results. Recent examples are the two MER rovers that carry a panoramic stereo camera, an alpha-particle x-ray spectrometer for elemental composition, a Mössbauer spectrometer for the identification of iron-bearing minerals, a miniaturized thermal infrared spectrometer for mineralogy, and a rock-abrasion tool to allow the measurement of 'fresh' rock (Squyres et al., 2003). Curiosity has a laser-induced breakdown spectrometer, an x-ray diffraction and fluorescence instrument, a neutron spectrometer, an environmental monitor package, a hand-lens imager, and a radiation assessment detector (Grotzinger et al., 2012). Network landers have not been flown yet but are generally agreed to be important (compare also **Chapter 10.03**). These landers will operate the same payload simultaneously at several locations. Applications are seismology to explore the interiors of planets, meteorology to explore planetary atmospheres, electromagnetic induction studies, and planetary geodesy. The NASA InSight mission will place a geophysical station on the surface of Mars in 2016. ESA is preparing for the ExoMars mission with the Pasteur rover to be launched in 2018 that will search for traces of extinct and extant life.

The search for extrasolar planets mostly uses two technologies, transit observations and radial velocity measurements through Doppler shift observations (see **Chapter 10.21**). While the former uses the slight dimming of the light of the central star as the planet passes between the star and the observer, the latter uses the Doppler shift of light from the star induced by the gravity pull of the planet on the star. Radial velocity measurements are done with ground-based telescopes or interferometers (such as the HARPS near Geneva, Switzerland, http://www.eso.org/sci/facilities/lasilla/instruments/harps.html) and constrain the mass of the planet. Transits are observed using ground-based and space telescopes such as COROT and Kepler and measure the radius of the planet. A planet observed with both methods can be characterized by an average density. Newly detected planets are candidates until they are confirmed by back-up observation. Several thousand candidates have been identified (most of them by Kepler), and roughly two thousand are confirmed. Future missions (such as PLATO (ESA) and TESS (NASA)) aim to detect planets of Earth size and smaller size orbiting bright stars and to do spectroscopy.

10.01.4 Planet and Satellite Orbits and Rotation States

There are some interesting commonalities between the orbits and rotation states of the planets and their major satellites. The rotation of the terrestrial planets is discussed in detail in

Chapter 10.04. The orbits lie mostly in a plane that is defined by the Earth's orbital plane, termed the ecliptic, and that is close to the equatorial plane of the Sun. The formal definition of the ecliptic plane adopted by the IAU calls for the ecliptic pole to be the mean orbital angular momentum vector of the Earth–Moon barycenter.

The normals to the actual orbital planes have inclinations relative to the ecliptic normal that are small, only a few degrees. The dwarf planets differ with orbital inclinations between 11° and 47° (Table 1). Small bodies (asteroids and comets) also tend to have larger orbital inclinations. The rotation axes of most planets and major satellites are within a few tens of degrees to the vertical to their orbital planes. Notable exceptions are Uranus, whose rotation axis lies almost in its orbital plane; Venus, whose retrograde rotation can be expressed as an inclination of almost 180°; and Pluto, whose inclination is between those of Uranus and Venus. The reasons for these anomalies are unknown but are sometimes speculated to be attributable to impacts of planet-size bodies on the young planets during the late stage of accretion.

The rotation and revolution of most planets and moons are prograde, that is, counterclockwise if viewed from above the celestial north pole. Exceptions are Venus, Uranus, and Pluto whose retrograde rotations can also be described as inclinations of more than 90° of their rotation axes to their orbital plane normals. Another exception is Triton, whose retrograde orbital motion about Neptune is unusual. Triton, like Pluto, is speculated to have originated in the Kuiper Belt. It was likely captured into its retrograde orbit by Neptune (see Chapter 10.17). There are many more satellites with retrograde orbits. However, these irregular satellites are typically much smaller than Triton, objects in a size range of only a few kilometers to a few tens of kilometers. The known rotation periods of the major satellites and of most satellites in general are equal to their orbital periods. This is termed a 1:1 spin–orbit coupling and is believed to be the result of tidal evolution (see Chapters 10.17 and 10.18). Tidal evolution occurs because planets raise tides on their satellites just as the Moon raises tides on the Earth. The gravitational force of the planet then pulls on the tidal bulge of the satellite. This torque will reduce the rotation rate of the satellite – should it be greater than its average orbital angular speed – and the interaction will end with the observed 1:1 coupling. As a result, the satellites present their primaries with mostly the same face at all times.

The planet Mercury is particularly interesting with respect to spin–orbit coupling since it is the only planet that is in such a resonance state with the Sun. However, its coupling is not 1:1, but 3:2 (see Table 1). The reason for this somewhat odd ratio is agreed to lie with the unusually large eccentricity of Mercury's orbit that has a value of 0.2 (Colombo, 1965; Chapter 10.04). Values of eccentricity <0.1 are more typical of planets and major satellites; the dwarf planets and small bodies have larger eccentricities. The large eccentricity of Mercury's orbit causes significant differences to arise between the constant rotational angular velocity and the orbital angular velocity that varies along the orbit. The two angular velocities could only be exactly equal at all times if the orbit were perfectly circular with zero eccentricity. The orbital velocity on an eccentric orbit increases towards the perihelion (the point of the orbit closest to the Sun) and decreases towards the aphelion (the point farthest from the Sun). Mercury's 3:2 resonance causes the orbital angular velocity to be equal to the rotational angular velocity at perihelion. This minimizes the tidal torque on Mercury and stabilizes the resonance.

The orbital distances of the planets from the Sun roughly follow a law with the distance of one planet to the Sun being roughly twice the distance to its inner neighbor. This rule is called the *Titius–Bode law* and works with Jupiter and Mars only if the asteroid belt is counted as a planet. The origin of the law is little understood. Regular relations between orbital distances or periods (the latter two are coupled through Kepler's third law of orbital motion) are not unusual in the solar system, however. The most prominent example is the Laplace resonance between the innermost three major satellites of Jupiter: Io, Europa, and Ganymede. A comparison of the orbital periods in Table 1 shows that these are in the ratio of 1:2:4. The origin of the Laplace resonance and its stability is attributed to tidal interactions between Jupiter and its three resonant moons. The orbits expand in the resonance as rotational energy is transferred through tidal interaction from Jupiter to Io and passed on in part to Europa and Ganymede (see Chapters 10.08, 10.17, and 10.18). The age of the resonance and how it formed are unknown. From the amplitude of libration, it has been concluded that the resonance is relatively young. Thermal-orbital modeling, however, suggests an age of at least 2 billion years (Fischer and Spohn, 1990). The tidal interaction and, in particular, the dissipation of tidal energy are believed to be the cause of volcanic activity on Io and of a subsurface ocean on Europa (Chapters 10.08 and 10.18). This ocean is completely covered by icebergs and ice shields that may slowly move relative to each other, and the ocean may even harbor, or may have harbored, primitive forms of life (see Chapter 10.14). The mathematical theory of tidal evolution is outlined in an appendix to Chapter 10.17.

10.01.5 Composition and Interior Structure of Planets

The chemical composition of the solar system is mainly dominated by the composition of the Sun, which has 98.9% of the mass of all bodies in the system. Although the compositions of the planets are different from that of the Sun, they are related to the Sun's composition and reflect varying degrees of depletion in volatile elements (see articles in Volume 1 of the Treatise on Geochemistry, Davis (2004)). The Sun is mainly composed of hydrogen and helium but contains all the other elements found in the solar system in abundances that are characteristic of the so-called solar composition.

Hydrogen and helium are dominant in Jupiter and Saturn, which are massive enough to keep these elements in molecular form against their tendency to escape to space. The potential of a planet to keep a specific element is measured by its escape velocity, which is proportional to the square root of the ratio between the planet's mass and its radius and must be compared with the thermal speed of an element, which is inversely proportional to its molecular weight (see Chapter 10.13). Still, Jupiter and Saturn are depleted in both H and He with respect to the composition of the Sun (Chapter 10.16). In addition to H and He, Jupiter and Saturn contain substantial amounts of water (H_2O), ammonia (NH_3), and methane (CH_4),

compounds collectively known as the planetary ices because of their occurrence on the surfaces of the major planets' icy satellites. In the deep interiors of the giant planets, these compounds are to be found in their supercritical forms for which there is no difference between the gaseous and liquid states. In addition to H, He, and planetary ices, Jupiter and Saturn likely have central cores of rock and iron. Models of the interior structure of planets can be constructed from a sufficiently detailed knowledge of their *figures* and their *gravity fields*. (The gravity fields of the terrestrial planets are discussed in Chapter 10.05; interior structure models of the terrestrial planets in Chapter 10.02. Chapter 10.04 discusses the rotation of these planets and how the rotation depends on interior structure. The giant planets are discussed in Chapter 10.16, while interior structures of giant planet satellites are discussed in Chapter 10.18) The construction of interior structure models is a primary task of planetary physics. Rotation deforms the planets from spheres into prolate spheroids. The flattening of both the figure and the gravity field is dependent upon the variation of density with depth. Unfortunately, these data do not allow unique models. The most accurate method of exploring interior structure is *seismology* (Chapter 10.03), the study of sound waves that travel through the interiors of planets and of global oscillations. This method is the prime method for exploring the Earth's interior to provide models of the variation of density with depth and also laterally by a technique that is known as seismic tomography (see Volume 1 of the Treatise on Geophysics). Seismometers, the instruments to record the waves, have been placed on the Moon during the Apollo program, and the interpretations of the results have provided important insights even up to the recent discovery of a solid inner core in the Moon (Weber et al., 2011). The InSight mission currently build by NASA in cooperation with CNES and DLR will place a state-of-the-art very broad band seismometer on the surface of Mars in 2016. The acoustic coupling between the surface and the atmosphere offers chances to do seismology from the orbit using cameras, infrared mappers, and ionospheric sounders. This method offers particularly good chances for Venus where environmental conditions on the surface are exceedingly demanding. The method can also be applied at the giant planets to study free oscillations and constrain interior models with these.

Models of the giant planets suggest that the outer layers of both Jupiter and Saturn are mostly H and He. H is molecular at moderate pressures but becomes metallic at pressures larger than 170 GPa (Figure 9). This pressure is equivalent to depths of about 14 000 km in Jupiter and 27 000 km in Saturn (Chapter 10.16). The phase transformation occurs at greater depth in Saturn because the pressure increases at a smaller rate in this planet due to its smaller mass. The deeper interiors of both contain the denser ices possibly in a shell above the densest iron/rock core. The cores plus the ice shells are believed to have masses that are quite uncertain, up to 45 Earth masses and radii of a few Earth radii.

Uranus and Neptune are further depleted in H and He and consist mainly of ice and the rock/iron component in addition to some H and He. Their structure, as the interpretations of the gravity field reviewed in Chapter 10.16 suggest, is not as clearly layered as those of Jupiter and Saturn. Rather, there appears to be a gradual increase of density with depth accompanied by a gradual increase of the abundance of ice at the expense of H and He followed by a gradual increase of the abundance of rock/iron at the expense of ice.

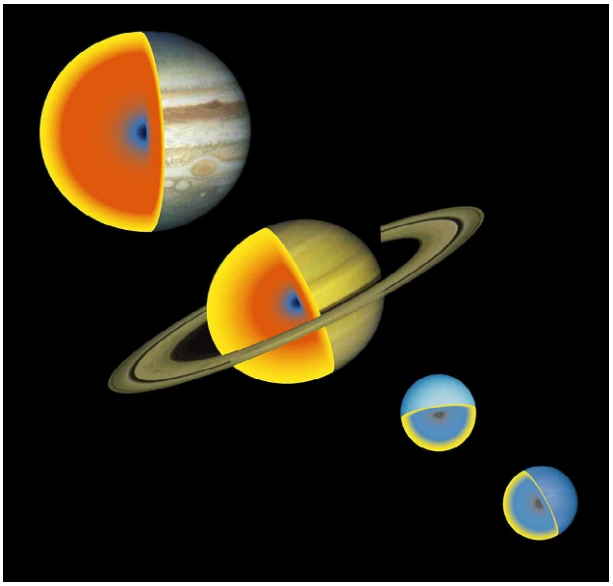


Figure 9 Interior structures of the giant planets. The yellow color indicates the molecular H and He shroud. Orange indicates the metallic H and He layers, blue is the range where the ice components are found, and the dark colors indicate the rock/iron cores. Courtesy of T. Guillot.

The major satellites of the outer solar system and Pluto also contain substantial amounts of ice, as their densities, around 2000 kg m^{-3}, suggest (see Table 1 and Chapters 10.18, 10.19, 10.20). Examples to the contrary, with substantially greater densities, are Io and Europa, the two innermost Galilean satellites of Jupiter. It is tempting to speculate that these two also started their evolution with a substantial abundance of ice. Io may have lost the ice completely and Europa to a large extent as a consequence of heating by impacts or as a consequence of tidal heating. Since Io and Europa are closer to Jupiter than Ganymede and Callisto, both the energy of impactors and the tidal dissipation rate will have been larger and may have caused the loss of water. But it is not clear that tidal heating could have provided the power needed to vaporize or partly vaporize ice layers on the two. However, it is also plausible that the temperature in the Jovian nebula from which the satellites formed was too hot at Io's and Europa's orbital distance to allow the accumulation of substantial ice shells. (The accretion of the satellites of Jupiter is discussed together with the formation of other satellites in Chapter 10.17) The Galileo mission has returned two-way Doppler ranging data that allow reasonable modeling of their interior structure (Figure 10). (See Chapter 10.18 for the icy satellites and Chapter 10.02 for Io. See also Chapter 10.23 for a brief discussion on the acquisition of radio science data.) Accordingly, Io has an iron-rich core of about half the satellite's radius and 20% of the satellite's mass with a rock shell above. Europa has a (water) ice shell about 150 km thick overlying a silicate rock shell and an iron-rich core. The nonuniqueness of interpretation of the data allows a wide range of sizes of these reservoirs, but the most likely radius values are 500–600 km for the core and 800–1000 km

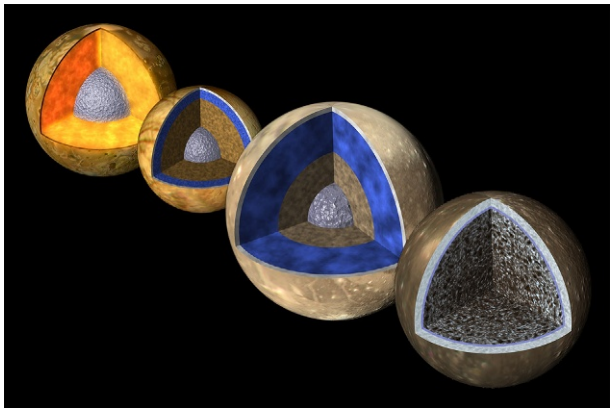


Figure 10 Interior structures of the Galilean satellites of Jupiter, Io, Europa, Ganymede, and Callisto (background to foreground).

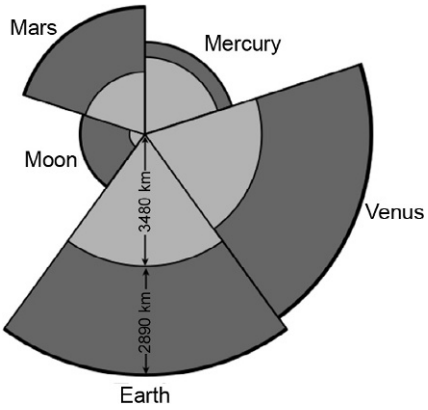


Figure 11 Basic interior structure of the terrestrial planets. To the first approximation, all terrestrial bodies possess a layered structure consisting of an innermost metallic core (light gray section), a silicate rocky mantle (dark gray layer), and a thin crust of volcanic origin, which is chemically distinct from the mantle (black layer not drawn to scale). While planetary radii are known precisely, core radii (light gray section) are not, apart from the case of the Earth's core, which consists of a liquid part surrounding a solid inner core (not drawn) with a radius of 1220 km. Reproduced from Tosi N, Breuer D, and Spohn T (2014) Evolution of planetary interiors. In: Spohn T, Breuer D, and Johnson TV (eds.) *Encyclopedia of the Solar System*. Amsterdam: Elsevier.

for the rock mantle. It is widely believed that a large part of the ice shell may actually be liquid allowing for an internal ocean. This is suggested by the results of the Galileo mission magnetic field measurements that indicate a field induced by Europa's motion in the magnetic field of Jupiter. Moreover, model calculations show that tidal dissipation in the ice may produce enough heat to keep an ocean underneath an ice lid a few tens of kilometers thick. It is also possible, depending on little known values of *rheology* parameters and thermal conductivity, that radioactive decay and tidal heating in the rock mantle may keep the ocean fluid (**Chapter 10.18**).

Ganymede has an ice shell about 800 km thick, a 900 km thick rock mantle shell, and a metallic core with about the same radius. The case for an iron-rich core is particularly appealing for Ganymede because the magnetometers on Galileo have detected a self-generated field on this satellite (see **Chapter 10.06**). The intrinsic magnetic field of Ganymede is most likely generated in an iron-rich core. The magnetic field data also suggest an induced component although the evidence for the induced field is much less strong. If the latter interpretation is correct, then Ganymede may also have an ocean of perhaps 100 km thickness at a depth of around 150 km. Since tidal heating is negligible at present, the heat that keeps the ocean molten must either be derived from radioactive decay in the rock mantle or from heat stored in the deep interior during earlier periods of strong tidal heating or even from accretion. In any case, the existence of the self-generated field is evidence enough for a molten core or, at least, a molten core shell.

Callisto is unusual among the Galilean satellites because the gravity data suggest that its interior is not completely differentiated. It is likely that there is an ice shell a few 100 km thick overlying a mixed interior of ice and rock also containing iron. It appears as if Callisto has traveled an evolutionary path different from that of Ganymede. The layering may have formed as a consequence of the slow unmixing of the ice and rock/iron components (Nagel et al., 2004). The model requires that Callisto has never been heated above the melting temperature of the ice. It is possible if not likely that the iron is no longer present in its metallic form but is oxidized to form magnetite and fayalite. The oxidation of the iron, if it occurred, would have precluded the formation of a metallic iron core. As for Europa and Ganymede, the magnetic field data suggest an induced field and an ocean for Callisto as well. Since Callisto is not in the Laplace resonance, tidal heating can be discarded as a heat source. The Cassini gravity data for the largest Saturnian satellite Titan suggest that it too is incompletely differentiated, thereby resembling Callisto, a satellite, the surface of which is so different.

The terrestrial planets (**Figure 11**) have ice in only modest concentrations and mostly consist of the rock/iron component. The rock/iron component in these planets, just as in Io, Europa, and Ganymede, is differentiated into a mostly iron core, a silicate rock mantle, and a crust consisting of the low-melting point components of rock (**Chapter 10.02**). The crust, a layer of some tens of kilometer thickness, forms by partial melting of the mantle and melt separation from the mantle. Crustal rock is still being produced on Earth by volcanic activity, but volcanic activity has mostly ceased early on Mars and Mercury (compare **Chapters 10.02**, **10.08**, and **10.09** and Toplis et al. (2013) for discussions on Mars).

Seismology has provided us with a detailed image of the interior structure of the Earth. It is expected that future missions will provide us with similar data for the other terrestrial planets (see **Chapter 10.03**). The most likely candidate here is Mars and the Moon, for which mission scenarios with seismological networks have been repeatedly studied. The seismological data for the Earth show that there are phase transformations mostly at moderate pressures and depths and chemical discontinuities. The existence of these layers and their thicknesses and depths will vary among the planets. For instance, it is likely that Mercury's mantle will not have major phase-change layers simply because the pressure in this small planet does not reach the levels that it reaches within the Earth, but it may be chemically layered as is the Moon. Although the latter is even smaller than Mercury, the seismic data available suggest a layered structure.

The case of the Moon illustrates the importance of data archiving. Only recent progress in seismology has provided the tools to more fully exploit the Apollo data and has demonstrated the layering of the lunar interior including a liquid outer and a solid inner core (Weber et al., 2011). The layering of even small bodies may be due to the very early evolution of the planets. There is evidence that the planets were once covered by what is called a magma ocean, rock molten by the dissipation of the energies of infalling planetesimals during the late stages of accretion. The Moon possibly became chemically layered as the magma ocean fractionally crystallized. The Earth may have escaped that layering because vigorous convection mixed the products of fractional crystallization. The vigor of convection depends on the size of the body or layer undergoing convection.

10.01.6 Surfaces and Atmospheres

The surfaces of most planets and major satellites have been observed optically, in the visible wavelength range of the electromagnetic spectrum, by cameras onboard spacecraft. The gaseous giant planets and subgiants show the top layers of their cloudy atmospheres. Since these bodies lack solid surfaces, their surface radius has been defined to be the radius at which the atmospheric pressure is 10^5 Pa. Mercury and the major satellites except Titan show their solid surfaces. These bodies are lacking substantial atmospheres. The ability of a planet or satellite to keep an atmosphere depends on the planet's gravity (compare **Chapter 10.13**). Equating the kinetic energy of a molecule to its potential energy defines the escape velocity that is independent of the mass of the particle. Atoms, ions, or molecules move at their thermal speed – which obeys a Maxwell–Boltzmann distribution and is proportional to the square root of the ratio of the temperature to the particle mass – but can be accelerated by other processes much beyond their thermal speeds. Among these processes are impacts, photoionization, and pickup by the solar wind, hydrodynamic effects, and sputtering. The presence of a magnetic field will help to keep an atmosphere by blocking the solar wind from eroding the atmosphere. Escape due to the thermal speed (at the long end of the Maxwell–Boltzmann distribution) exceeding the escape velocity is called the Jeans escape and is most relevant for light species such as hydrogen and helium. The Jeans escape can be used for a systematic discussion in terms of planetary mass, radius, and temperature. It explains why Mercury with its comparatively small mass and high surface temperature is prone to losing an atmosphere and why massive Jupiter can bind hydrogen and helium, but it cannot explain why Titan has a substantial atmosphere, while Ganymede, similar in mass, radius, and temperature, has none. The difference may lie with Titan being rich in ammonia from which the mostly nitrogen atmosphere may have formed by photodissociation (Coustenis and Taylor, 1999 and **Chapter 10.18**). It is also possible that the difference between Titan and Ganymede lies with their early differentiation and outgassing histories. The compositions of the atmospheres of the terrestrial planets suggest that any solar-type primordial atmosphere has been replaced by a secondary atmosphere that is the result of outgassing and perhaps cometary impacts (e.g., Pepin, 2006).

The surface temperatures of the planets are mostly determined by the solar radiation and therefore decrease with increasing distance from the Sun. The actual surface temperature will depend on the reflectivity or *albedo* in the visible wavelength of the electromagnetic spectrum, on the emissivity in the infrared, and on the thermal and optical properties of the atmosphere. A useful quantity, the *effective temperature*, can be calculated assuming thermal equilibrium of the surface with the solar radiation neglecting the effects of the atmosphere. The effective temperature decreases from 445 K at Mercury to 50 K for Neptune (**Table 1**). In case of an atmosphere, the energy balance is more complicated (**Chapter 10.13**). Visible light passes through the atmosphere and reaches the surface where it is in part reflected and in part absorbed and reradiated. The infrared radiation emanating from the surface is in part absorbed by CO_2 and/or H_2O in the atmosphere and raises the atmosphere temperature. This effect, the so-called greenhouse effect, is particularly important for Venus (see also Bengtson et al. (2013)) due to the large amount of CO_2 in its atmosphere. (Other minor constituents contribute to the greenhouse effect.) Here, the surface temperature is about 740 K, about twice the value of the effective temperature. It is possible that Mars had a much denser atmosphere in its early evolution that allowed for higher temperatures through the greenhouse effect and liquid water. This would have helped to develop life on this planet (**Chapter 10.14**).

The solid surface of the Earth is partly covered by clouds in the atmosphere and partly visible depending on the extent and the pattern of the cloud cover. Venus' surface, however, is entirely covered by an optically dense atmosphere and the same is true for the Saturnian satellite Titan. Venus' atmosphere is mostly CO_2 and the atmospheric pressure is 9.2 MPa. The pressure at the bottom of Titan's atmosphere is 150 kPa. The orange-colored haze that hides Titan's surface from view is most likely smog produced by the photodissociation of methane. Venus' surface has been explored by radar observation. Titan is also explored by radar through the Cassini mission, which, at the time of this writing, is orbiting the planet and has repeatedly flown past Titan at a distance of about 1000 km. Radar and infrared mapping data have revealed a variety of morphological surface features such as lakes, dunes, and perhaps ice volcanoes suggesting a complex story of climate, geology, and atmosphere/surface interactions (e.g., Jaumann et al., 2009). On 14 January 2005, the European Huygens probe onboard Cassini successfully landed on Titan and returned the first images from its surface. The images reveal a complex surface morphology and suggest rivers and perhaps oceans (**Figure 12**).

The surface of Mars is visible most of the time. The main component of the Martian atmosphere is CO_2 but the atmospheric pressure is much smaller than Venus', of the order of 600 Pa. However, the surface may be hidden during times of global dust storms, which occur repeatedly on the timescales of a few years. Ancient river beds, outflow channels, and erosion features have been taken as evidence that the Martian atmosphere was more massive in the past and the climate was wetter and warmer (see **Chapter 10.11** for a discussion of water and ice on terrestrial planetary surfaces, in particular on Mars, and **Chapter 10.12** for a discussion of atmosphere–surface interactions). However, this view is debated and

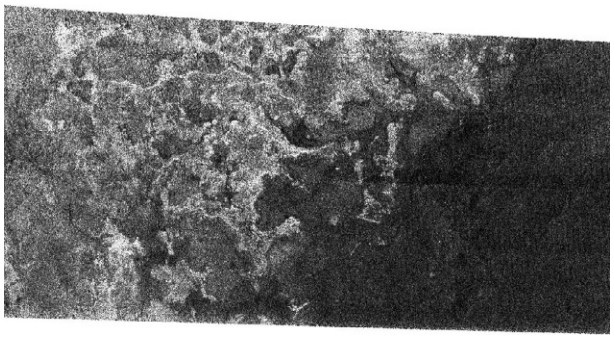


Figure 12 Synthetic aperture radar image of the surface of Titan. The bright, rough region on the left side of the image seems to be topographically high terrain that is cut by channels and bays. The boundary of the bright (rough) region and the dark (smooth) region appears to be a shoreline. The patterns in the dark area indicate that it may once have been flooded, with the liquid having at least partially receded. The image is 175 km high and 330 km wide and is located at 66 degrees south latitude, 356 degrees west longitude in the southern hemisphere of Titan.

alternative scenarios are discussed (**Chapter 10.11**). The findings of the Mars Exploration and Curiosity rovers clearly show that standing bodies of water on Mars were once present in which water-bearing minerals like jarosite precipitated. Water is a prerequisite for a planet to be habitable.

The pressure on Mars is close to the triple-point pressure of water (611 Pa). Thus, there is little room for liquid water except for low-lying regions such as the bottom of impact basins like Argyre and Hellas. However, melt water may exist for some time metastably on Mars and longer as its surface freezes over and it is covered by ice (see also articles in Tokano (2005) and references therein). The widely accepted early disappearance of the magnetic field of Mars (see Schubert et al. (2000) for a dissenting view) may have been partly responsible for the escape of the early Martian atmosphere (see Dehant et al. (2007) for a recent review).

The atmosphere of the Earth differs from those of its neighbors. The atmospheric pressure of 0.1 MPa is in between that of the latter two, and the main components are N_2, O_2, and H_2O.

The outer surface of Uranus is bland, greenish in color, and mostly featureless. The greenish color is attributed to methane and high-altitude photochemical smog. At the other extreme is Jupiter, whose surface features a large number of bands or stripes largely parallel to the equator interspersed with spots and vortices. Particularly remarkable is the giant red spot, a vortex that covers about 1/10 of the planetary disk (see **Figure 1**). These features point to a highly dynamic atmosphere. The dynamics is dominated by the rotation as witnessed by the band structure, but the vortices and spots show that these become unstable at various scale lengths. The brownish to reddish color of some of the features is attributed to ammonium hydrogen sulfide (NH_4SH), while the whitish colors are attributed to ammonium. Saturn's atmosphere is much calmer than Jupiter's, as the surface patterns suggest. These patterns are stripes similar to those in the Jovian atmosphere, but the spots and vortices are missing although wind speeds are extremely high. Neptune is similar to Uranus in that its surface is bluish in color, which is attributed to methane. It resembles Saturn in its stripe pattern. But in addition, there are a few vortices and a large spot that resemble features on Jupiter (the atmosphere dynamics of the giant planets is discussed in **Chapter 10.16**).

The solid surfaces of the terrestrial planets have some common features. Most prominent among these are craters that occur on a very wide range of sizes (compare **Chapter 10.10**). Craters are believed to be remnants of the early evolution of the planets and satellites when the young surfaces were bombarded by planetesimals during the late stage of the accretion process. This phase is often termed the phase of heavy bombardment. The distribution and density of craters on planetary surfaces are an important indication of the age of the surface or parts thereof. The older a surface is, the higher the density of impact craters. For instance, Mars shows a dichotomy in surface age (and topography) between its northern lowland and southern highland hemispheres (see **Figure 8**). The age difference on average is about 1 billion years. Some of the most cratered and therefore oldest surfaces in the solar system are found on Mercury, the Earth's Moon, and the Jovian satellite Callisto. There are surfaces that are saturated with craters, implying that for every new crater formed by an impact, one existing crater will be – on average – destroyed. Geologists and planetary scientists use the term 'exogenic dynamics' to characterize the processes modifying the surface by impacts. Many impact craters on terrestrial planets with atmospheres have been partly or completely destroyed by erosion. Some have been buried by extensive resurfacing in the early history of the planet. On icy satellites, long-term relaxation of the ice may cause craters to flatten out.

Planetary surfaces are also modified by erosion (**Chapter 10.12**) and by processes related to 'endogenic dynamics,' that is, by processes that originate from within the planet (see **Chapters 10.08** and **10.09**). The most prominent of the latter processes is volcanism. Volcanism has shaped at least parts of the surfaces of the terrestrial planets, and silicate volcanism is the one important element of crustal growth. Prominent volcanic features are the giant Tharsis dome on Mars, the island arcs on the Earth, and the Maria on the Moon. The surface of Venus is dotted with volcanoes and plains. As we will discuss further in the succeeding text, endogenic activity scales with the mass of a planet. This partly explains why the activity on smaller planets usually dies off earlier than on the larger planets.

A special case is the Jovian satellite Io (see **Chapter 10.08** and articles in Spencer and Lopes (2006)). This satellite is similar in mass to the Moon, yet it is the most volcanically active planetary body in the solar system. Volcanic features cover its surface but impact craters have not been detected. This shows that the surface is very young and is permanently renewed. The surface is to a large extent covered by allotropes of sulfur and by sulfur dioxide. These deposits cause the yellow to whitish color of the surface. The dark spots are likely volcanic vents and sulfur lakes. Sulfur deposits are often found associated with volcanic structures on the Earth. The high surface temperatures of up to 1600 K indeed suggest silicate volcanism. The reason for the unusual activity on Io lies with an unusual heat source: Io is flexed by tides raised by its massive primary Jupiter. The deformation energy that is dissipated as heat is sufficient to make it more volcanically active

than the largest terrestrial planet, the Earth. A further measure of the enormous energy that is dissipated in Io's interior is the surface heat flow of 2–3 W m^{-2}, 20–30 times larger than the surface heat flow of the Earth. Io's heat flow is rivaled only by Jupiter, which radiates about 5.4 Wm^{-2}. The heat flow through the southern polar cap of Enceladus is also exceptionally high, however (**Figure 13**).

The Earth features a style of endogenic activity, plate tectonics, that, according to our present knowledge, is unique to the Earth (chapters in Volumes 6 and 7 and **Chapters 10.08** and **10.09**). Plate tectonics involves the continuous production of basaltic crust along volcanically active linear ridges at the bottom of the oceans. Prominent examples are the Mid-Atlantic Ridge and the East Pacific Rise. It also involves subduction of this crust underneath island arcs and continental margins. Prominent examples are the islands of Japan for the former and the western continental margin of South America for the latter. These subduction zones are the loci of most of the seismic activity of the planet. Both processes cause the Earth's surface to be divided into seven major plates that drift across the surface. Plate tectonics is driven by extremely slowly circulating convection currents in its deep interior. Although convection is not unique to the Earth but is expected to occur in the silicate rock shells of the other terrestrial planets as well, this feature is usually hidden underneath a thick stagnant lid (see **Section 10.01.7** and **Chapters 10.08** and **10.09**). The reason why convection in the Earth extends to the surface and includes the crust at the bottom of the oceans is not entirely known. It is speculated that this is caused by the presence of water and there may even be links between plate tectonics and life (e.g., Höning et al., 2014; Sleep et al., 2012). In any case, plate tectonics causes the surface of the ocean basins to be completely renewed on a timescale of a few hundred million years. Plate tectonics together with erosion even incorporates recycling of the more stable continents. Material denudated by erosion from the continents is transported to the ocean basins where it is incorporated into the plate tectonics cycle. That loss of continental material is balanced by the production of continental rock through volcanic activity. The cycling of material between the interior and the atmosphere is an important element of the carbon silicate and other elemental cycles that help stabilize the climate of the planet (cf. **Chapter 10.13**). It has been speculated that Mars and Venus went through phases of plate tectonics very early in their histories (see **Chapters 10.08** and **10.09**). However, while this is a possibility, it remains a speculation. Although plate tectonics seems to be unique to the Earth, there are other processes of crust recycling. These involve subcrustal erosion as is assumed for, for example, Io and the foundering of the surface as has been speculated to have happened on Venus at the sites of some coronae Ghail (2002) even proposed a mechanism of "buouyant plate tectonics" for present day Venus.

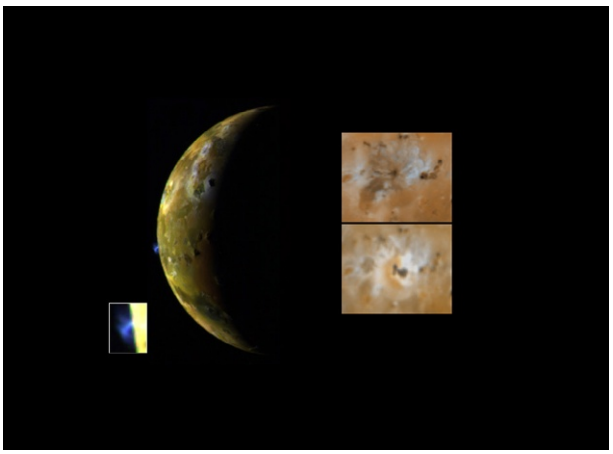


Figure 13 Volcanic plume on Io imaged by the Galileo spacecraft. The blue-colored plume extends to about 100 km above the surface. The blue color is consistent with the presence of sulfur dioxide gas and sulfur dioxide snow condensing as the volcanic gas in the plume expands and cools. The images on the right show a comparison of changes seen near the volcano Ra Patera since Voyager in 1979 (top) and Galileo in 2001 (bottom). An area of 40 000 km^2 was newly covered with volcanic material between the two observations. NASA/JPL.

The best studied example of a single-plate planet is Mars. Three instruments on board spacecraft have recently allowed major progress for detailed studies of the tectonics of this planet (**Chapter 10.09**). The first is the laser altimeter MOLA on Mars Global Surveyor, the second is the HRSC camera on Mars Express, and the third is the HiRISE camera on the Mars Reconnaissance Orbiter. While MOLA has stopped operating as an altimeter in 2001 after covering the entire surface due to the finite lifetime of its laser pump diodes, HRSC has covered almost the entire planet at a resolution of 10 m px^{-1} at the time of this writing and HiRISE has supplemented this data set with images down to 30 cm px^{-1}. One major achievement by these instruments is the establishment of an accurate geodetic network of surface features.

The topography of Mars exhibits a clear dichotomy that divides the surface into a southern highland hemisphere as heavily cratered as the lunar highlands and rising several thousands of meters above the zero level and a northern lowland hemisphere that lies well below the datum. The origin of the dichotomy is ascribed variously either to long-wavelength mantle convection, to impacts, to postaccretional core formation sweeping up most of the crustal material into one large protocontinental mass, or to combinations thereof (see discussion in **Chapters 10.08** and **10.09**). The highlands cover about 60% of the planet, including almost all of the southern hemisphere. The surface has survived since the heavy bombardment prior to 3.8 billion years ago with only minor modifications and thus records the early history of the planet.

Venus whose surface has been mapped by the Magellan and Venus Express missions with radar and in the infrared is another well-studied single-plate planet. Its hypsometric curve, however, has a single peak rather than having double peaks as those of Mars' and the Earth's curve. Venus has no distinct volcanic center such as Mars but a prominent rift system, the Beta, Atla, and Themis Regiones. Volcanic features include a large number of comparatively small shield volcanoes that are apparently randomly distributed, volcanic plains, and about 500 coronae – circular large-scale volcanotectonic features. It is widely accepted that most coronae form above upwelling plumes (Smrekar and Stofan, 1999). The comparative paucity of impact craters on the surface has been used to suggest that Venus was resurfaced about 500–700 million years ago by a volcanic event of global scale (Schaber et al., 1992;

McKinnon et al., 1997). This event was followed – as the model suggests – by almost no volcanic activity. The early conclusions from the cratering record have been challenged by, for example, Hauck et al. (1998) and Campbell (1999). According to these authors, the cratering record allows a variety of interpretations in terms of volcanic resurfacing including a global decrease in time in the rate of volcanic activity.

10.01.7 Energy Balance and Evolution

Although the evolution of the planets seems to have followed some common general lines, there are significant differences among individual planets. The evolution of the terrestrial planets and moons is discussed in **Chapter 10.08** and of the icy satellites in **Chapter 10.18**. The evolution of the giant planets is discussed in **Chapter 10.16**. The early evolution of all planets has been dominated by impacts from the debris left over from planet formation (**Chapter 10.10**). Moreover, isotopic evidence suggests that the planets differentiated by iron-core formation early within a few tens of million years (e.g., Kleine et al., 2002). Of course, the evidence for Mars rests with the very well-founded assumption that the so-called SNC meteorites are in fact rocks from Mars. It is further believed that the planets started hot, heated by the energy deposited in the interior during accretion and by heat dissipated upon differentiation – the terrestrial planets heated to near-melting temperatures of their rock components and to temperatures well above the iron-melting temperatures. The planets then cooled from this initial hot state and the cooling drove their evolution. It is possible, for instance, to explain the present rate of infrared radiation from Jupiter simply by cooling. This is not possible for Saturn, for which an additional heat source is required. The continuing gravitational settling of helium may provide enough energy to explain the observed present luminosity of this planet. The energy balances of most other planets and satellites seem to invoke both cooling and heat generated by the decay of radioactive elements, in particular of uranium, thorium, and potassium (the isotope ^{40}K), which generates 4–5 pW kg^{-1} of rock. A further possibility is tidal heating, but this mechanism seems to be relevant at present only for Io and, perhaps, Europa and Enceladus.

It should be stressed that the energy balances of the planets are not very well known. A crucial quantity, the surface heat flow – or the intrinsic luminosity – is known for the giant planets, which radiate enough energy in comparison with the solar insolation incident on their surface that their luminosity can be measured from orbit. This is also true for the heat flow from the volcanic hot spots on Io, but there is a major uncertainty concerning the flow through the remainder of the surface. Nevertheless, the enormous luminosity of this satellite of at least 1.5 W m^{-2} is remarkable. Another remarkable value is the roughly 6 GW (Spencer et al., 2006) radiated from the southern hemisphere of Enceladus. On the terrestrial planets and the other satellites, as far as we know them today, the surface heat flow must be measured *in situ* (compare **Chapter 6.05** by Jaupart). This measurement requires a borehole deep enough to avoid the influence of the daily and seasonal temperature variations of the atmosphere. This difficult measurement has been done at many locations on the Earth and at two sites on the Moon. The InSight mission scheduled for launch to Mars in 2016 will attempt a surface heat flow measurement. In **Table 1**, we list the specific luminosities or surface heat flows for the planets and satellites for which the heat flows and luminosities have been measured. Since heat production depends on mass, we have divided the heat flows by the masses. A comparison of the entries in **Table 1** shows that the Earth and the Moon are within a factor of two of the specific radioactive heat production rate of rock suggesting that about half of the heat flow can be attributed to cooling and half to heat production. This conclusion is known to hold for the Earth even on the basis of more specific data and more careful energy balances (compare **Chapters 9.08** and **9.09**). An educated prediction of the heat flows for Mars, the other terrestrial planets, and the major satellites could be based on the aforementioned observation. The specific luminosity of Uranus also is close to the radiogenic heat production rate per unit mass of rock given in the preceding text. It is likely that Uranus has a substantial rock core in which that heat is generated.

Neptune, Io, and Europa stand out in comparison with the terrestrial planets. In the case of Neptune, the extra heat flow has been attributed to cooling and whole-planetary contraction. For Io, the extraordinary large heat flow is with little doubt due to tidal heating although ohmic dissipation of energy carried by electric currents between Io and Jupiter is sometimes quoted. The value for Europa in **Table 1** is highly uncertain and has been derived from indirect observation of tectonic surface features (e.g., Ruiz, 2005). If the estimate were correct, then tidal heating would be the best explanation for its value. However, equilibrium models of tidal heating arrive at values smaller by about one order of magnitude (cf. **Chapter 10.18**).

The dynamic or geologic, chemical, and magnetic evolutions of the planets are mainly governed by their thermal evolution. Planets can be regarded as heat engines that convert heat into mechanical work and magnetic field energy. Mechanical work will be performed by the building of volcanic structures and mountain belts, for example, or during the movement of plates over the Earth's surface (cf. Volumes 6 and 7). The agent for these processes is convective heat transport, a mode of heat transport that involves movement of hot material to cold surfaces and cold material to hot surfaces (see, e.g., Schubert et al. (2001) and **Chapter 7.02**). Convection is driven by sufficiently great temperature differences across a layer or planetary shell and occurs if the layer is unstable against convective overturn. Convection is possible even in the solid rock mantles of the planets because of the long timescales that are involved. On geologic timescales of tens to hundreds of millions of years, rock behaves as a very viscous fluid. The convection flows at speeds of centimeters per year still transfer more heat than conduction. This is in part due to the enormous masses involved in the flow and in part due to the very low thermal conductivity of rock. There is even a thermostat principle at work in the silicate and ice shells (the Tozer principle) (Tozer (1967)) since the viscosities of these solids are strongly temperature-dependent: If the flow is too

slow to transport the heat, the temperature will increase until the flow is strong enough to transfer the heat. If the temperature is so high such that the convective heat transfer rate largely outbalances the heat production rate or the heat flow into the layer from below, then the temperature will decrease, viscosity will increase, and convective vigor will decrease. Convection is extensively discussed in Volume 7. The thermal evolution of the Earth is discussed in Volume 9. Chemical differentiation is mostly a consequence of melting and the density difference between melt and solid that will again lead to movement. A particularly good example for the chemical differentiation of the terrestrial planets that occurs in addition to core formation is the growth of the crust through volcanic activity (see Chapter 10.08). However, secondary crustal growth likely involves much longer timescales than core formation.

The material in the crust is derived from partial melting of the mantle. Partial melting is possible when a material of complex chemistry lacks a simple melting temperature. Rather, in materials such as rock (and ice of complex chemistry), there will be a temperature termed the solidus temperature at which the material begins to melt. The melt will consist of that component of the entire assemblage that has the lowest individual melting point. The remaining solid will be depleted in that component. As the temperature rises, other components begin to melt until, finally, at the liquidus temperature, the whole assemblage will be molten. In the rock component of planets, the component to most easily melt is basalt. Once a basalt liquid forms in the interior of a planet, it will tend to rise to the surface by virtue of its lower density than the density of the remaining rock. In a water–ammonium ice, the low-melting component will be a water–ammonium mixture of a particular composition. In most cases, the melt is likely to be produced within the top few 100 km of a planet by a process termed pressure-release melting. Pressure-release melting occurs when the melting point gradient is steeper than the average temperature gradient. Relatively hot uprising convection currents will emanate from depths where the temperature is below the melting temperature. Upon rising to the surface adiabatically (i.e., with no or little heat exchange), the flow will hit a depth where the solidus temperature is exceeded and partial melting will ensue. Because melt is usually much more compressible than solid, there will be a pressure and a depth below which melt will no longer be buoyant to rise to the surface but may actually sink towards the deeper interior. Melting in the interior of planets requires a heat source. This source is mainly the heat that is stored in the planets during accretion and the heat that is generated by the decay of radiogenic elements. Both are finite reservoirs that will not be replenished. That is why endogenic activity decreases with time.

Since melting occurs when hot material from the deep interior rises in convection currents, the surface distribution of volcanism is speculated to give an indication of the planform of the convection underneath. The timing of the volcanic activity can be used as a guide for assessing the time evolution of the convection. For instance, the differing surfaces of Mars and Venus may have recorded differing planforms of mantle convection and flow history. The dominance of Tharsis on Mars (compare Figure 8) suggests that there is or that there once was a giant superplume, a very large upwelling underneath this volcanic dome. The geologic evidence suggests that Tharsis formed early but that volcanic activity has been ongoing to the recent past (cf. Chapter 10.08). This long-term stability of Tharsis as the major center of volcanic activity on Mars is puzzling, the more so since model calculations suggest that the superplume's lifetime should be not much more than one billion years. Although Tharsis apparently formed early, the photogeologic evidence also suggests that volcanic activity on Mars started globally and retreated to Tharsis over time and may then have decreased in vigor. On Venus, there is an indication that the volcanic activity has been global even recently (on geologic timescales, recently involves the past 10–100 million years). There is further (albeit debated) evidence for a global volcanic resurfacing event a few hundred million years ago. This suggests that the thermal history of a planet may have involved episodes of more and of less activity that most likely were linked to episodes of greater and less vigor of convection. Volcanism may also have influences on a hydrosphere and atmosphere by degassing the deep interior of volatile elements. At the time of strong volcanism, an increase of atmospheric gases or water on the planet's surface is expected.

There is a fundamental difference between the evolutions of the Earth and most other terrestrial planets and major satellites that is related to the occurrence of plate tectonics on the Earth (discussed in Chapters 10.08 and 10.09). Since the viscosity or, more generally, the rheology of rock is strongly temperature-dependent and since the surface temperatures of these planets are much lower than the melting temperatures, it follows that there must be outer layers that are comparatively stiff. These layers are termed the lithospheres. Usually, the lithospheres are connected lids that are pierced here and there by volcanic vents. Convection currents flow underneath the lithosphere and deliver heat to the base of the lid through which it is then transferred to the surface by heat conduction. Since these lids are stagnant, this form of tectonics is often termed stagnant-lid tectonics. On Earth, however, the lithosphere is broken into seven major plates that move relative to each other driven by the convective flow underneath. Volcanic activity occurs along some plate margins and results in the growth of plates. Other plate margins are destructive and are loci where plates are forced under their own weight to subduct into the mantle. There is no convincing evidence that plate tectonics occurs on any other planet or satellite although it has been speculated that it may have occurred on early Mars and on the Jovian satellites Europa and Ganymede. There is evidence in the images of the surface units of lateral movement on these two satellites. The Jovian satellite Io does not seem to be undergoing plate tectonics. However, its present resurfacing rate, which is larger than the resurfacing rate of Earth, calls for some recycling of Io's crust volcanic material with the underlying mantle. This recycling probably occurs through delamination of the base of the crust. Delamination of the base of the crust may have been or may be operative on other planets such as Venus (Figures 14 and 15).

The interior evolution of the gaseous planets is even less well constrained. It must be assumed that the dynamics of their atmospheres is related to the vigor of convection underneath, but the solar insolation also matters. Thus, models of the evolution are mostly constrained by their luminosities, but as we have seen in the preceding text, these leave enough puzzles

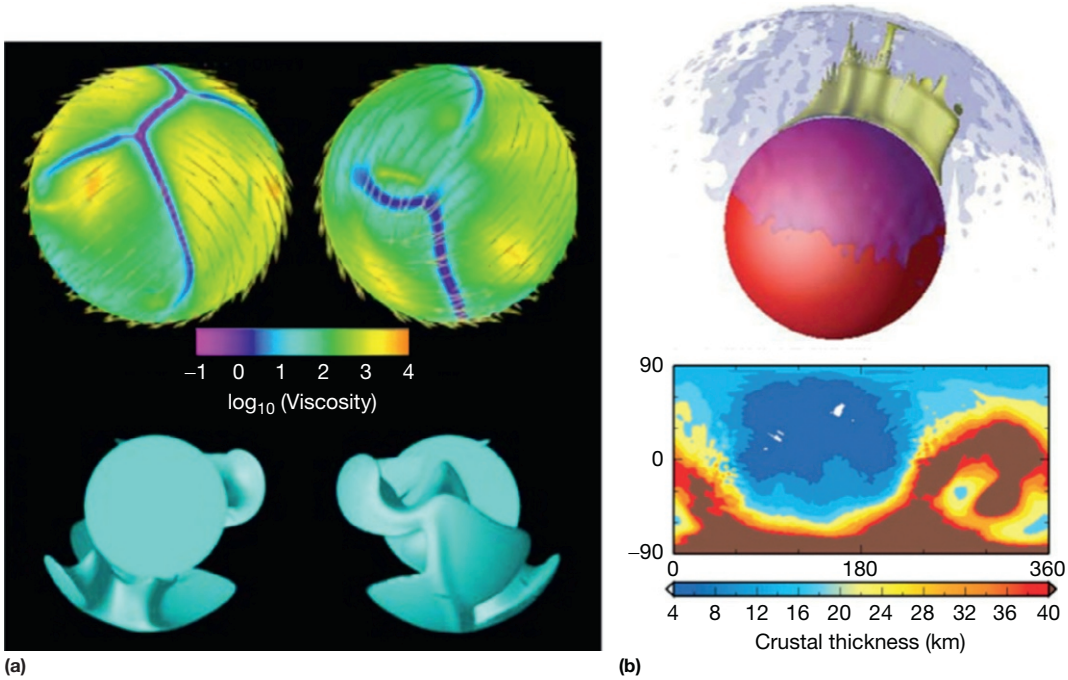


Figure 14 (a) Numerical simulations of mantle convection with plate tectonics in a three-dimensional spherical shell. The top line shows two views of the viscosity field at the surface. Rigid plates moving with uniform velocity are separated by both divergent and convergent narrow boundaries with low viscosity at which plates form and are recycled into the mantle. The bottom line shows temperature isosurfaces representing the subduction of cold downwellings (after van Heck H and Tackley PJ (2008) Planforms of self-consistently generated plate tectonics in 3-D spherical geometry. *Geophysical Research Letters* 35: L19312, http://dx.doi.org/10.1029/2008GL035190). (b) Numerical simulation of the formation of Mars' crustal dichotomy by mantle convection. The top panel shows a hot upwelling plume (yellow) rising from the core–mantle boundary (red) able to generate a large region of partial melt (light blue). Migration of the melt towards the surface leads to the formation of a crust (**Section 10.01.7**) whose thickness in the southern hemisphere is much greater than in the northern one (bottom panel) (after Šramek O and Zhong S (2012) Martian crustal dichotomy and Tharsis formation by partial melting coupled to early plume migration. *Journal of Geophysical Research* 117(E01005), http://dx.doi.org/10.1029/2011JE003867).

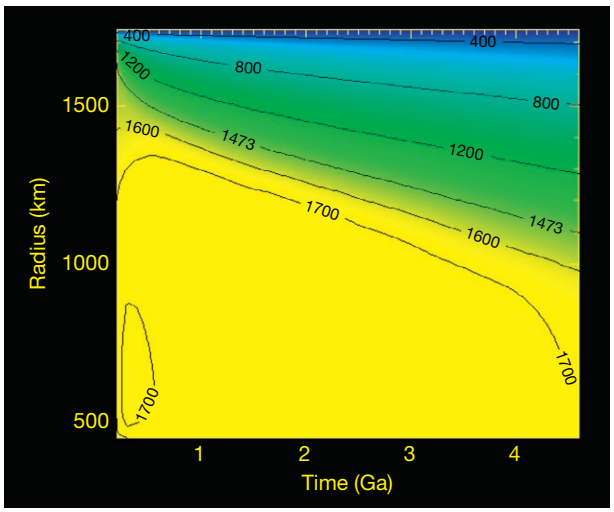


Figure 15 Growth of a stagnant lid in the lunar mantle. The growth of a cold, thermally conductive lid on top of the lunar mantle is shown as resulted from convection model calculations by Konrad and Spohn (1997).

to be solved as witnessed by the differences between the luminosities of Neptune and Uranus.

10.01.8 Magnetic Fields and Field Generation

Mercury, Earth, and the giant planets have largely dipolar magnetic fields that are produced by dynamo action in their interiors. Of the major satellites, only Ganymede is known to produce a magnetic field. The Galileo and Cassini data suggest that neither the other Galilean satellites of Jupiter – in particular not dynamic Io – nor the big Saturnian satellite Titan have self-generated fields. The magnetic fields of these planets and moons are described and discussed in **Chapter 10.06**. The physics of planetary dynamos is discussed in **Chapter 10.07**. **Chapter 10.08** discusses the generation of the magnetic fields of terrestrial planets in the context of their thermal evolution, and **Chapter 10.16** the generation of magnetic fields in the giant planets. **Chapter 10.18**, finally, discusses the magnetic fields of outer solar system satellites. Planetary magnetism is also extensively reviewed and discussed in Christensen et al. (2011).

The icy Jovian satellites Europa, Ganymede, and Callisto are surrounded by magnetic fields that vary along with their movement through the magnetosphere of Jupiter. These fields are, therefore, interpreted to be induced in electrically conducting layers in the satellites' interiors as the satellites orbit Jupiter and are subject to a time-varying magnetic flux. The geometry and the strengths of these fields suggest that the conducting layers are at depths of some tens to a few hundred kilometers and are most likely salty oceans. No evidence for an induced field at Titan has been reported. Thermal considerations suggest that these oceans are feasible (**Chapter 10.18**). Ganymede is agreed to have an induced field on top of its permanent magnetic field.

Self-sustained magnetic fields are generally thought to be enigmatic to planets, part of their interior evolution during which thermal (and potential energy) is converted into mechanical work and magnetic field energy. The Mars Global Surveyor data have confirmed this hypothesis by showing that Mars has a remanently magnetized crust. This crust must have been magnetized during an epoch when Mars was generating its own magnetic field. Interestingly, it is mostly the oldest crust units that are magnetized. It has long been speculated that Mars, Venus, and the Earth's Moon once produced magnetic fields by dynamo action in their cores (e.g., Stevenson et al., 1983). The observation of magnetized rock on the surface of Mars now confirms this general notion (**Figure 16**).

The dynamo mechanism (e.g., **Chapter 10.07** but see also **Chapters 10.08** and **10.16** and chapters in Volume 8, in particular **Chapter 8.03**) invoked to explain magnetic field generation is similar among the planets although the source regions differ between the terrestrial planets and major satellites, the giant planets, and the subgiants or ice giants, Uranus and Neptune. Required is an electrically conducting and fluid region that undergoes turbulent flow driven by thermal or compositional buoyancy. If a magnetic field exists, for instance, the external magnetic field of the Sun or a primary planet, a field will be induced in the source region. The flow will distort the field lines and thereby generate magnetic field energy by induction. If the flow satisfies certain conditions, the interference between the generated field and the preexisting field can be constructive, powering and maintaining a magnetic field against dissipative losses.

For the terrestrial planets and the major satellites, the candidate source regions are the fluid metallic cores or outer core shells. Because of their low viscosities (around 1 Pa), these regions are often unstable with respect to thermal convection. The stability is generally accepted to depend on the rate of heat removal from the core by the mantle convection flow. If the latter is too low (below about 10 mW m^{-2} but depending on pressure), the heat flow in the mantle can be balanced by heat conduction in the core, and the core will be stably stratified. Dynamo action is then not possible. If the heat flow into the mantle is larger than the critical value, convection in the core must be invoked to balance the heat removal rate. The situation becomes more favorable for convection in the core if the core begins to freeze and to grow a solid inner core. First, latent heat liberated upon inner core growth will help to power the dynamo. Second, the core may contain light alloying elements such as sulfur and/or oxygen. The light alloying elements expelled from the solid inner core can drive (chemical) convection and a dynamo very effectively because this dynamo will not be subject to a Carnot efficiency factor that by the second law of thermodynamics limits the efficiency at which a heat engine can do work. In this model of the chemical dynamo, the melting curve is always steeper than the core (adiabatic) temperature profile. It is believed that the cooling of the core and growth of the inner core drive the Earth's core dynamo.

A new class of chemical dynamos described in **Chapter 10.08** has been discussed in recent years in particular for terrestrial planets and satellites smaller than the Earth. Examples are the Earth's Moon, Ganymede, Mercury, but possibly even Mars.

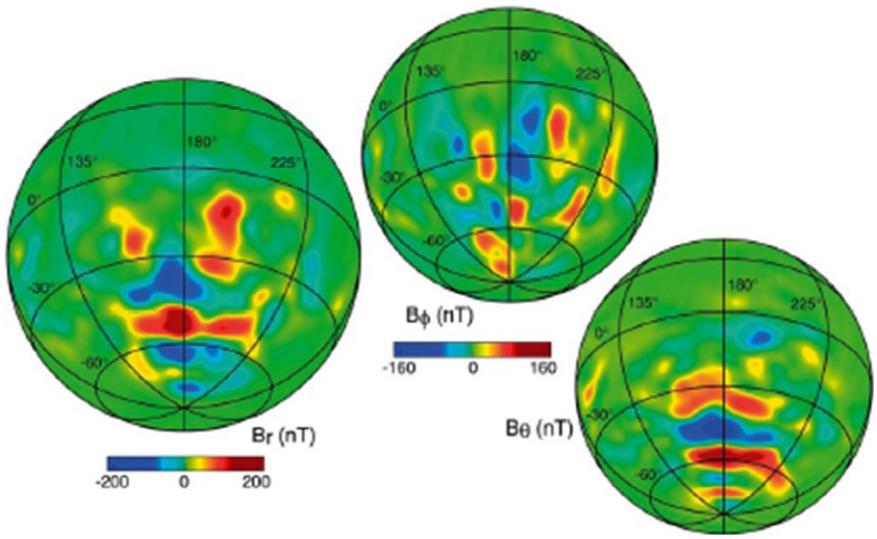


Figure 16 Orthographic projections of the three components of the magnetic field in spherical coordinates (r, radius; Θ, latitude; and Φ, longitude) onto the surface of Mars at a nominal mapping altitude of 400 km. Reproduced from Connerney JEP, Acuna MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111: 1–32.

The new model is based on the observation that the iron-rich core alloy melting relations may be more complicated than in the older models where the application of the Lindeman melting law warranted that the melting was steeper than the adiabat. In particular, it is possible that the melting curve is less steep than the adiabat and the gradient may even be negative. In both cases, the core will not freeze from the center outward, but iron snow will form in the outermost parts of the core. As the iron snow sinks into the core, it will form a partially molten layer in which solid iron and melt are in equilibrium underlain by a layer of melt enriched in iron with respect to the average composition of the core alloy. The latter layer will be gravitationally instable and drive chemical convection and possibly a dynamo. As the core cools, the partial melt layer (with neutral buoyancy) will grow and eventually the dynamo becomes frustrated. The model may be able to explain why the Moon can have a solid inner core at present without generating a magnetic field.

Dynamo action in a terrestrial planet thus depends on the efficiency of mantle convection, on the composition of the core, and on the core material phase diagram. Plate tectonics is very effective at cooling the deep interior of a planet because the cold plates sink deep into the mantle. It is therefore conceivable that the Earth's core has been cooling to temperatures below the liquidus of its core alloy. Venus, a planet of similar size, appears to have been lacking that efficient cooling mechanism and appears to be cooling by convection underneath a stagnant lid. Consequently, the heat flow from the core has become subcritical over time, and the core may have not been cooling enough to reach liquidus temperatures. As a consequence of this and the general decline of mantle convective vigor, the core became stably stratified and a possible early dynamo ceased to operate. A similar scenario is possible for Mars, but the iron snow model may provide an alternative explanation for an early dynamo that ceased to operate. It would be interesting to see whether or not Mars has a solid inner core. There is little hope to find remanently magnetized crust on Venus should landing missions be able to overcome the forbidding operating conditions on this planet. The surface temperature of around 450 °C is above the temperatures at which candidate minerals become remanently magnetized. The Moon has a remanently magnetized crust and may be another good candidate for the iron snow model. Alternative explanations for the recorded remanently magnetized crust units invoke plasma clouds generated by major impacts. Some of the strongest magnetic anomalies (albeit weak in comparison to Mars) are suspiciously located at the antipodes of major impact basins. Mercury differs from both the Moon and Mars because of its extraordinarily large core and thin rocky mantle. The thin mantle should be quite effective, as model calculations suggest, at removing core heat and driving a dynamo that produces a magnetic field, albeit weak. The weakness of the field has been discussed to be difficult to explain with conventional dynamo theory (compare **Chapter 10.07**).

Dynamo action in the gaseous giant planets (**Chapters 10.06**, **10.07**, and **10.16**) most likely happens in the regions where hydrogen becomes metallic due to the extraordinary large pressure. The magnetic fields of these planets are most likely powered by the energy that has been stored in the planets during accretion and that is removed by convection. The transition to metallic hydrogen occurs at a depth of roughly 15 000 km in Jupiter and almost twice as deep in Saturn. The larger depth of dynamo action may partly explain why Saturn's field is much more ideal in terms of a dipole than Jupiter's. In the source regions, the fields are likely to be very complex and not at all similar to a dipolar planetary field. The field in the source region can be thought of as a superposition of many multipolar fields of various amplitudes. Distance from the source region affects the higher-order multipolar fields much more strongly than the lower-order fields. If n is the order ($n=2$ for a dipole), then the strength of the field components decreases with radial distance r from the source regions as r^{-n}. Thus, a planet with a comparatively deep source region should feature a more ideal dipolar field as compared with one where the source region lies less deep. This does not explain entirely the characteristics of the Saturnian magnetic field for which screening by helium rain has been additionally invoked.

The pressure in the interiors of the subgiants Uranus and Neptune is not sufficient to cause a transition to metallic hydrogen. Rather, it is believed that there are ionic oceans at relatively shallow depths in these planets. In these ionic oceans, the magnetic fields can be generated. The fields of the subgiants have been measured only by a single flyby each and are therefore not well known. Nevertheless, it appears that these fields are very complex in their topologies. This fits in nicely with the idea of the fields being generated at relatively shallow depths.

10.01.9 Origin of the Solar System

The origin of the solar system is not covered in this book. The reader is referred to Volume 1 of the Treatise on Geochemistry (Holland and Turekian, 2003). A most recent overview can be found in Chambers and Halliday (2014), and more detailed recent reviews in the Protostars and Planets V (Reipurth et al., 2007) and VI collection of articles (Beuther et al., 2014). The formation of satellites is discussed in **Chapter 10.17**.

The basic elements of the theory of the origin of the solar system are similar to the Kant–Laplace hypothesis of formation from a gaseous nebula, but the details are debated. It is widely accepted that the nebula collapsed to form a central mass concentration – the protosun – surrounded by a spreading thin gaseous disk. The most widely held view postulates that temperature in the inner part of the nebula soon became low enough – about 1500 K – to allow the condensation of silicate and iron grains. In the outer solar system, beyond about 5 AU, temperature became so small as to allow condensation of ice phases. Within the first few million years, these solid grains agglomerated to form bigger grains that, in turn, agglomerated to form even bigger ones. The cascading scenario of ever bigger and fewer planetesimals led to the formation of solid bodies the size and mass of the terrestrial planets in the inner solar system and a few tens of Earth masses in the outer solar system. In the outer solar system, these solid protoplanets became the cores of the giant planets. Growth of the solid protoplanets slowed to almost a standstill as the solid matter became exhausted, at least in the feeding zone of the protoplanets that is determined by the competition between a protoplanet's gravity and the gravitational pull of the Sun. At about one Earth mass, a protoplanet begins to

accrete gas onto its surface. The heat dissipated during the influx of planetesimals was radiated from the surface of the forming gaseous envelope. After the influx of solid matter ceased, the protogiant planets began to cool and to contract, thereby increasing its potential to accrete more gas. The growing mass led to even more accretion and runaway growth was set in place. Runaway accretion came to a halt when a gap formed in the nebula around the planet. This gap can form as a consequence of the competing gravities of the protoplanet and the Sun and as a consequence of the finite viscosity of the gas.

A relatively new feature of solar system formation models is that giant planets can move inward (and outward) through interaction with the nebula, thereby perturbing the orbits of smaller planetesimals and protoplanets. The popular Nice model (named after the city where its basic elements were developed; Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005) includes the effects of resonances between the orbits of the giant planets and offers explanations for the relative small mass of Mars along with late heavy bombardment and the formation of the asteroid belt.

The general model offers an explanation of the grand features of the solar system: The inner planets are solid and of refractory composition because the temperature in the inner nebula favored the condensation of refractory phases. The inner planets are small because the feeding zone was smaller with less mass of solid particles. The bigger of the inner planets – the Earth and Venus – have atmospheres with masses as expected although the present atmospheres are not likely to be the primordial ones, which were lost and replaced by degassing of the interiors. In the outer solar system, we find planets with cores of some ten Earth masses and massive gaseous envelopes. In addition, we find wealths of satellite systems witnessing accretion in orbit around the growing protogiants. There are a number of features that can be explained by fine-tuning the theory, for instance, the differences between Jupiter and Saturn and Uranus and Neptune but this will not change the grand picture.

Giant impacts, a mechanism of great importance for some bodies, are the final events of the accretion scenario. Giant impacts are thought to have caused the formation of the Moon (e.g., Canup, 2004; Galimov and Krivtsov, 2012) and may be responsible for the high density of Mercury. A giant impact is a collision of almost grown protoplanets. For instance, it is thought that the Moon formed after a Mars-sized protoplanet hit the proto-Earth. The outer layers of the Earth vaporized during the impact, and the Moon formed from the condensed vapor cloud. Among the arguments for the giant impact hypothesis for the formation of the Moon is the geochemical closeness of the bulk Moon to the Earth's mantle. Mercury may have suffered a similar giant impact that removed a substantial part of the original planetary mantle (e.g., Benz et al., 2008). This may explain why present Mercury has a comparatively big core and a thin mantle and thus an extraordinary large average density. A Moon did not form around Mercury because of the closeness to the Sun. Giant impacts may also affect the rotation, and it has been proposed that Venus' retrograde rotation may thus be explained. Similar explanations have been brought forward for Uranus.

10.01.10 Concluding Remarks

With this overview, I have tried to take the reader on a tour through the solar system and provide the stage for the detailed articles in the remainder of the volume. The planets are largely of solar composition but differ in their depletion in volatile elements. The degree of depletion increases with decreasing mass and with decreasing distance from the Sun with Jupiter being closest in composition to the Sun and with the terrestrial planets being mostly depleted in volatile elements. The planets are mostly internally differentiated with the heavy elements tending to be found near the center and the most volatile elements near the surface. The gaseous planets present the observer with their top layers of their envelopes, while most of the smaller planets present their solid surfaces. Relicts of very early impacts characterize these surfaces to varying degrees. But these surfaces also reveal the signs of volcanic and tectonic activity and erosion. In general, the smaller the solid-surface planet, the more characterized is its surface by early impacts. The volcanic activity and the magnetic fields of those planets that have self-generated fields are due to convection currents in their interiors driven by cooling and by heat generated by radioactive decay or by the dissipation of tidal energy. The source regions of the magnetic fields are metallic or, in the cases of Uranus and Neptune, ionic fluids. Life has had a chance to develop on Earth but possibly also on Mars and Europa. Finding evidence for life outside Earth is a challenge for the future. The bio–geo system of Earth may be a self-regulating system that sets this planet apart from its sister planets. Its position in the solar system, together with a moderate greenhouse effect in its atmosphere, has provided the planet with habitable temperatures and with an atmospheric pressure allowing for liquid water. Life plays an important role here by removing CO_2 from the atmosphere and moderating the greenhouse effect. Water, on the other hand, may be instrumental for plate tectonics since its effect on the rheology may pave the way for surface plates to subduct to the deep interior, thereby cooling the core. The cooling of the core is speculated to be instrumental for magnetic field generation. The magnetic field is a feature that the Earth's brother and sister planets Mars and Venus lack. The magnetic field protects the environment, and thereby life, from radiation. It will be important to see whether or not life developed in the less friendly environments of Mars or even Europa or early Venus.

The future in scientific planetary exploration on the one hand certainly lies with systematic remote sensing with long-term orbiters. These missions will create data volumes that require either long-term operation and communication or significantly increased data acquisition and transmission rates. On the other hand, *in situ* exploration with rovers and planes and networks of surface stations with long operational lifetimes to observe the atmospheres and the interiors will be important. A further challenge for the future will be the return of samples from planets and small bodies. The sophistication of terrestrial laboratories cannot simply be transferred to other planetary surfaces given the scarcity of resources such as mass and power on spacecraft. The technological challenge remains miniaturization. Although many planetary worlds are forbidding to human exploration and require the operation of robots

(think of Venus), it is probably true that the human explorer cannot be topped in its ability for *in situ* scientific study and operation of hardware.

Closer to home, the challenges for planetary geophysics and planetary science in general lie with numerical modeling and laboratory studies. For instance, the rheology of planetary matter is not sufficiently well understood as are the properties of materials in the deep interiors of the Earth and Venus, let alone Jupiter and the other giant planets. Numerical modeling has advanced largely in step with the advancement of computer power, but for geophysical modeling of, for example, the dynamo and mantle convection, the achievable parameter ranges are still in many cases far from the realistic ones.

Exploration has started to look beyond the solar system for other planetary systems. About two thousand confirmed extrasolar planets have been discovered to date in the mass and radius ranges from 10 Jupiter masses down to Earth size and even Mars size. Of course, interest is directed towards Earth-like planets in the habitable zone around Sun-like stars in other solar systems! These planets are beyond the reach of present observational tools, and space agencies are planning to launch space-based telescopes that will be more capable such as the planned ESA PLATO mission and the NASA TESS mission. Also, the James Webb Space Telescope will provide unprecedented opportunities.

But the questions are the following: Is the Earth typical? May other solar systems have planets with thick ice covers such as the icy moons of Jupiter, but bigger? Is it conceivable that there are planets of about the Earth's size with primordial, Jupiter-like atmospheres (Stevenson, 2004)? From a theoretical point of view, these worlds are possible (a statement that may simply reflect our ignorance). In the icy planets, radiogenic heat may support oceans and it is conceivable that these oceans harbor life. Could life even exist in gaseous envelopes?

References

Acuña MH, et al. (1999) Global distribution of crustal magnetism discovered by the mars global surveyor MAG/ER experiment. *Science* 284: 790–793.

Bagenal F, Dowling TE, and McKinnon WB (eds.) (2004) *Jupiter*. Cambridge: Cambridge University Press.

Basri G and Brown ME (2006) Planetesimals to brown dwarfs: What is a planet? *Annual Review of Earth and Planetary Sciences* 34: 193–216.

Bengtson L, Bonnet RM, Grinspoon D, Koumoutsaris S, Lebonnois S, and Titov D (eds.) (2013) *Towards understanding the climate of Venus*. ISSI Scientific Report, Vol. 11, Heidelberg: Springer, 185 pp.

Benz W, Anic A, Horner J, and Whitby JA (2008) The origin of Mercury. In: Balogh A, Ksonfomality L, and Steiger R (eds.) *Mercury. Space Science Series of ISSI*, vol. 26, pp. 7–20. New York: Springer.

Beuther H, Kiessen RS, Dullemond CP, and Henning T (eds.) (2014) *Protostars and Planets VI*. Tucson: University of Arizona Press, 952 pp.

Burke J (2014) Planetary exploration missions. In: Spohn T, Breuer D, and Johnson TV (eds.) *Encyclopedia of the Solar System*, p. 1205–1222. Amsterdam: Elsevier.

Campbell BA (1999) Surface formation rates and impact crater densities on Venus. *Journal of Geophysical Research* 104: 21,951–21,955.

Canup RM (2004) Dynamics of lunar formation. *Annual Review of Astronomy and Astrophysics* 42: 441–475.

Chambers JE and Halliday AN (2014) The origin of the solar system. In: Spohn T, Breuer D, and Johnson TV (eds.) *Encyclopedia of the Solar System*, p. 29–54. Amsterdam: Elsevier.

Christensen U, Balogh A, Breuer D, and Glassmeier KH (eds.) (2011) *Planetary Magnetism, ISSI Space Science Series*, vol. 33. Berlin: Springer, 668 pp.

Colombo G (1965) Rotational period of the planets Mercury. *Nature* 208: 575.

Coustenis A and Taylor F (1999) *Titan: The Earth-like Moon*. London: World Scientific Publishing, 344 pp.

Davis AM (ed.) (2004) *Treatise on Geochemistry*, vol. 1. Amsterdam: Elsevier.

Dehant V, Lammer H, Kulikov YuN, et al. (2007) The planetary magnetic dynamo effect on atmospheric protection of early Earth and Mars. *Space Science Reviews* 129: 279–300.

dePater I and Lissauer JJ (2010) *Planetary Sciences*, 2nd edn. Cambridge: Cambridge University Press, 647 pp.

Fischer HJ and Spohn T (1990) Thermal-orbital histories of visco-elastic models of Io. *Icarus* 83: 39–65.

Galimov EM and Krivtsov AM (2012) *Origin of the Moon. New Concept: Geochemistry and Dynamics*. Berlin: W. de Gruyter, 179 pp.

Gargaud M, Amils R, and Cernicharo J, et al. (eds.) (2011) *Encyclopedia of Astrobiology*, vol. I—III. Heidelberg: Springer.

Ghail R (2002) Structure and tectonics of south-east Tethis Regio. *Journal of Geophysical Research* 107: 1–7.

Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.

Grotzinger JP, Crisp J, Vasavada AR, et al. (2012) Mars Science Laboratory mission and science investigation. *Space Science Reviews* 170: 5–56. http://dx.doi.org/10.1007/s11214-012-9892-2.

Hauck SA II, Phillips R, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103: 13,635–13,642.

Höning D, Hansen-Goos D, Airo A, and Spohn T (2014) Biotic vs. abiotic Earth: A model for mantle hydration and continental coverage. *Planetary and Space Science* 98: 5–13, http://dx.doi.org.10.1016/j.pss.2013.10.004.

Holland HD and Turekian KD (eds. in chief) (2003) *Treatise on Geochemistry*, 2nd edn., vol 1–9. Amsterdam: Elsevier.

Iess L, Stevenson DJ, Parisi M, et al. (2014) The gravity field and interior structure of Enceladus. *Science* 344: 78–80.

Jaumann R, Kirk R, Lorenz R, et al. (2009) Geology and surface processes on Titan. In: Brown R, Lebreton JP, and Waite H (eds.) *Titan from Cassini-Huygens*, pp. 75–140. New York: Springer.

Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.

Konrad W and Spohn T (1997) Thermal history of the moon: Implications for an early core dynamo and post-accretional magmatism. *Advances in Space Research* 19: 1511–1521.

Lebreton J-P, Witasse O, Sollazzo C, et al. (2005) An overview of the descent and landing of the Huygens probe on Titan. *Nature* 438: 758–764.

Lodders K and Fegley B Jr. (1998) *The Planetary Scientist's Companion*. New York: Oxford University Press, 371 pp.

McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*. Tucson: University of Arizona Press.

Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.

Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution and differentiation of Callisto. *Icarus* 169: 402–412.

Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on Mars revealed by the High Resolution Stereo Camera. *Nature* 432(2004): 971–979.

Pepin RO (2006) Atmospheres on the terrestrial planets: Clues to origin and evolution. *Earth and Planetary Science Letters* 252: 1–14.

Porco CC, Baker E, Barbara J, et al. (2005) Cassini imaging science: Initial results on phoebe and Iapetus. *Science* 307: 1237–1242.

Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.

Reipurth B, Jewitt D, and Keil K (eds.) (2007) *Protostars and Planets V*. Tuscon: Arizona University Press.

Ruiz J (2005) The heat flow of Europa. *Icarus* 177: 438–446.

Russel C and Raymond C (2011) *The Dawn Mission to Minor Planets 4 Vesta and 1 Ceres*. Heidelberg: Springer.

Saal AE, Hauri EH, Cascio ML, Van Orman JA, Rutherford MC, and Cooper RF (2008) Volatile content of lunar volcanic glasses and the presence of water in the Moon's interior. *Nature* 454: 192–195.

Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus: What are they telling us? *Journal of Geophysical Research* 97: 13,257–13,301.

Schubert G, Olsen P, and Turcotte DL (2001) *Mantle convection in the Earth and Planets*. Cambridge: Cambridge University Press, 956 pp.
Schubert G, Russell CT, and Moore WB (2000) Timing of the Martian dynamo. *Nature* 408: 666–667.
Shirley JH and Fairbridge RW (eds.) (1997) *Encyclopedia of Planetary Sciences*. London: Chapman and Hall, 990 pp.
Sleep NH, Bird DK, and Pope E (2012) Paleontology of Earth's mantle. *Annual Review of Earth and Planetary Sciences* 40: 277–300.
Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.
Smrekar SE and Stofan ER (1999) Evidence for delamination of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.
Sotin C, Jaumann R, Buratti BJ, et al. (2005) Release of volatiles from a possible cryovolcano from near-infrared imaging of Titan. *Nature* 435: 786–789.
Spencer JR and Lopes R (2006) *Io After Galileo: A New View of Jupiter's Volcanic Moon*. Berlin: Springer, 342 pp.
Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hotspot. *Science* 311: 1401–1405.
Spohn T, Breuer D, and Johnson TV (eds.) (2014) *Encyclopedia of the Solar System*, 3rd edn. Elsevier: Amsterdam, 1311 pp.
Squyres SW, Arvidson RE, Baumgartner ET, et al. (2003) Athena Mars rover science investigation. *Journal of Geophysical Research* 108: 8062. http://dx.doi.org/10.1029/2003JE002121.
Squyres SW and Knoll AH (eds.) (2005) *Sedimentary geology at Meridiani Planum, Mars*, *Earth and Planetary Science Letters* 240: 1–190.
Stevenson DJ (2004) Planetary diversity. *Physics Today* 57: 43–48, Special Issue Planetary Diversity.
Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism of the terrestrial planets. *Icarus* 54: 466–489.
Tokano T (ed.) (2005) *Water on Mars and Life*. Heidelberg, NY: Springer, 331 pp.
Toplis M, Bell J, Chassefiere E, Sotin C, Spohn T, and Blanc M (eds.) (2013) *ISSI Space Science Series*, vol. 44. *Quantifying the Martian Geochemical Reservoirs*. Heidelberg: Springer, 364 pp.
Tozer DC (1967) Towards a theory of thermal convection in the mantle. In: Gaskell TF (ed.) *The Earth's Mantle*, pp. 325–353. London: Academic Press.
Trümper J (ed.) (2009) *Solar System, Landolt-Börnstein Numerical Data and Functional Relationships*, *New Series VI/4B*. Berlin: Springer, 628 pp.
Tsiganis K, Gomes R, Morbidelli A, and Levison H-F (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Nature* 435: 459–461.
Weber RC, Lin PY, Garnero EJ, Williams Q, and Lognonne P (2011) Seismic detection of the lunar core. *Science* 331: 309–312.

10.02 Interior Structure, Composition, and Mineralogy of the Terrestrial Planets

F Sohl, German Aerospace Center (DLR), Berlin, Germany
G Schubert, University of California, Los Angeles, CA, USA

© 2015 Elsevier B.V. All rights reserved.

10.02.1 Introduction 24
10.02.2 Observational Methods 25
10.02.2.1 Geodesy 25
10.02.2.2 Rotation and Tides 26
10.02.2.3 Gravity and Topography 26
10.02.2.4 Magnetic Fields 27
10.02.2.5 Electromagnetics 27
10.02.2.6 Seismology 28
10.02.2.7 Surface Geology and Composition 28
10.02.2.8 Material Properties 29
10.02.3 Interior Structure and Composition 29
10.02.3.1 Two- and Three-layer Structural Models 29
10.02.3.2 Depth-Dependent Structural Models 30
10.02.3.2.1 Governing equations 31
10.02.3.2.2 Boundary conditions 31
10.02.3.2.3 Equation of state 32
10.02.4 Earth as a Type Example of a Terrestrial Planet 33
10.02.4.1 General 33
10.02.4.2 Interior Structure 33
10.02.4.3 Composition 34
10.02.4.4 Mineralogy 34
10.02.5 Vesta 34
10.02.6 The Moon 35
10.02.6.1 General 35
10.02.6.2 Interior Structure 35
10.02.6.3 Composition 37
10.02.6.4 Mineralogy 38
10.02.6.5 Future Exploration 40
10.02.7 Mercury 40
10.02.7.1 General 40
10.02.7.2 Interior Structure 42
10.02.7.3 Composition 43
10.02.7.4 Magnetic Field 43
10.02.7.5 Future Exploration 44
10.02.8 Mars 45
10.02.8.1 General 46
10.02.8.2 Interior Structure 46
10.02.8.3 Composition 49
10.02.8.4 Mineralogy 51
10.02.8.5 Martian Seismicity 52
10.02.8.6 Future Exploration 52
10.02.9 Venus 52
10.02.9.1 General 52
10.02.9.2 Interior Structure 53
10.02.9.3 Composition 54
10.02.9.4 Tectonism 54
10.02.9.5 Dynamics 54
10.02.10 Solid Exoplanets 55
10.02.11 Summary and Outlook 55
Acknowledgments 56
References 56

10.02.1 Introduction

The inner planets Mercury, Venus, Earth, and Mars and terrestrial-type bodies like the Moon and some of the outer planet satellites are mainly composed of silicate rock and metals like iron. They are characterized by relatively small masses and radii and large densities in comparison with the giant planets of the outer solar system. Measurements of rotation and gravitational and magnetic fields indicate that the interiors of these bodies, like the interior of the Earth, are chemically and rheologically layered and subdivided into iron-rich cores, silicate mantles, and rocky crusts derived from partial melting of the mantles. It is generally believed that the internal differentiation of the terrestrial planets and major satellites took place early in their histories only shortly after accretion from colliding planetesimals (Kleine et al., 2002). Even the asteroid 4 Vesta, an accretional remnant left over from that period, reveals a variegated surface with characteristic spectral features hinting at a strongly differentiated interior (Drake, 2001; Ghosh and Mcsween, 1998; Keil, 2002; Kleine et al., 2002; Righter and Drake, 1996, 1997).

The investigation of planetary interiors is among the most important scientific objectives of interplanetary space missions. The state of knowledge of planetary differentiation provides important clues on the general understanding of the internal structure, thermochemical evolution, and bulk composition of the terrestrial planets and terrestrial-type bodies like the Moon and a number of outer planet satellites. Many large-scale planetary processes are controlled by the internal structure of these bodies. Surface geology and tectonic features are mainly affected by mechanisms that dominate the transport of internal heat from the interior to the surface. The existence of self-sustained and/or induced magnetic fields requires reservoirs of electrically conducting fluids at some depth, providing additional constraints on the present thermal state of these bodies. Since a fluid layer within a planetary body mechanically decouples the deep interior from its outer portion, the propagation of seismic waves, the way in which a planet or satellite responds to tides, and the rotational state in terms of spin-axis precession rate, obliquity, forced libration amplitude, and nutations are strongly affected by the physical state of its interior.

Models of the interior structure of a number of terrestrial-type planetary bodies are shown in **Figure 1**. These models are based on theoretical considerations that are strongly constrained by Earth-based and remote-sensing observations, in situ measurements, and laboratory studies of planetary materials and meteorites. Interior structure models aim at calculating (1) the volumes and masses of major chemical reservoirs (crust, mantle, and core) that contribute to the bulk composition, (2) the depths to chemical and rheological discontinuities and mineral phase boundaries, and (3) depth variations of pressure, temperature, density, and composition. In the absence of seismological data, the most important

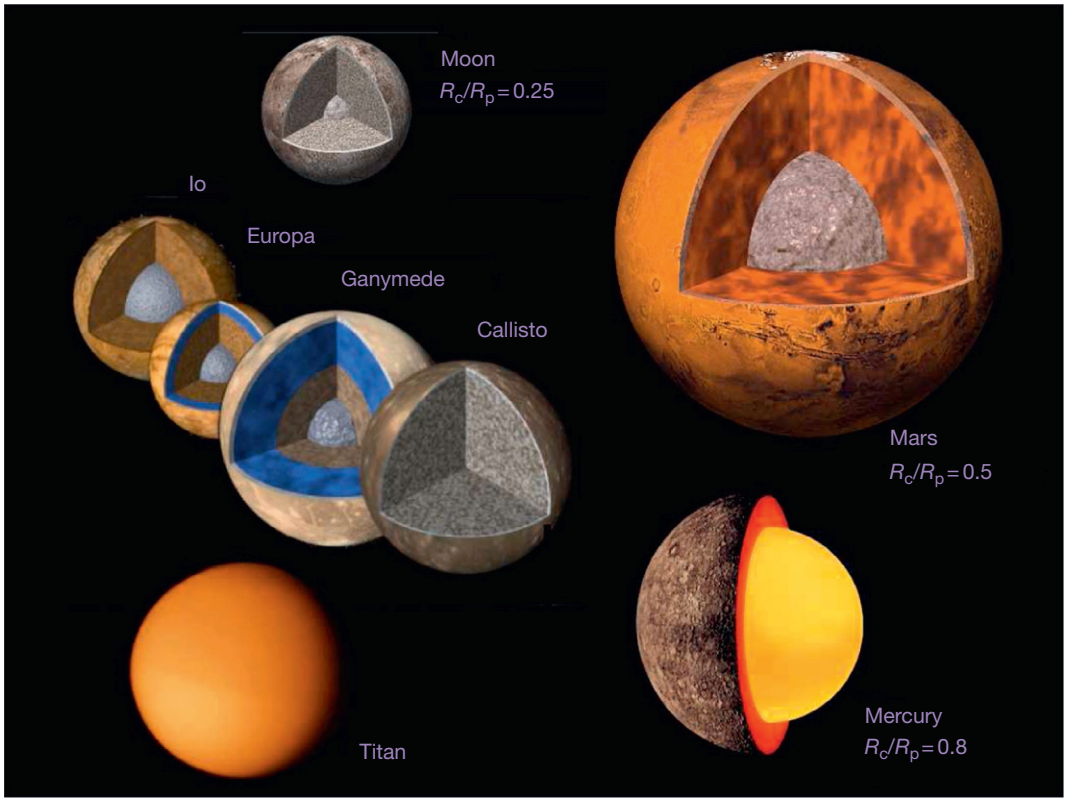


Figure 1 Cutaway views of the interiors of a number of terrestrial-type bodies. R_c/R_p denotes the core radius of a planet or satellite relative to its total radius. Reproduced from © NASA/JPL/RPIF, Calvin J. Hamilton; Strom RG (1987) *Mercury*: *The Elusive Planet*. Washington, DC, London: Smithsonian Institution Press, with permission.

parameter that permits a rough estimate of how the interior is composed is the average or mean density.

The mean density is affected by bulk composition, thermal state, self-compression, and pressure-induced mineral phase transitions caused by the weight of overlying layers. In lieu of seismological observations, the determination of the axial moment of inertia (MoI), usually by measurements of the spin-axis precession rate, augmented by low-degree gravity field data and possibly the magnitude of forced libration of the outer solid shell, provides the principal constraint on the concentration of mass toward the center of a planet or satellite. Large-scale gravity and topography data are also important in constraining internal mass distributions since the shapes of the physical surface and the external gravitational field are tied to the radial density distribution and compositionally and/or thermally induced lateral density heterogeneities.

This chapter is arranged as follows: In Section 10.02.2, those observational methods that provide constraints on the interior structure and composition of terrestrial planetary bodies are reviewed. Section 10.02.3 provides the mathematical background for the construction of spherically symmetrical density models and discusses thermodynamic quantities and associated equation-of-state parameters relating ambient material properties to local (pressure and temperature) material properties. Section 10.02.4 describes the interior of the Earth as a type example of a terrestrial planet and is followed by Sections 10.02.5–10.02.10, characterizing in some detail the interiors of the asteroid 4 Vesta, the Moon, Mercury, Mars, Venus, and solid exoplanets. Finally, a summary of this chapter is included in Section 10.02.11.

10.02.2 Observational Methods

Important observational constraints on planetary interiors are provided by astronomical and geodetic methods, rotational and tidal variations, relationships between gravity and topography, magnetic field observations, planetary seismology, interpretations of geologic surface features and compositional variations, and laboratory studies of planetary materials and meteorites.

10.02.2.1 Geodesy

In the absence of seismic data, bulk composition and interior structure of a planet or satellite are mainly constrained by its mean density and mean MoI. The mean density of a planetary body is calculated from its mass and volume. To that end, it is necessary to infer the size and global shape of a planet or satellite from images acquired telescopically using micrometric or photographic techniques or by cameras on board spacecraft. The global shape or figure of a body depends on its mass, size, rotational and tidal state, and material strength. Detailed topographic maps of Mercury, Venus, the Moon, and Mars have been obtained by using Earth-based radar observations (Harmon, 2007; Harmon et al., 2001, 2011; Margot et al., 1999; Muhleman et al., 1995), radio occultation data, and radar and laser altimetry from orbiting spacecraft (Rappaport et al., 1999; Smith et al., 1999a; Zuber et al., 1992, 1994, 2008, 2012). Though the masses of planets and satellites have been known from astronomical observations for a long time, the most current and accurate values are based on measurements of spacecraft that landed, orbited, or flew by.

The gravitational field of a planet or satellite is closely related to the distribution of mass inside the body. Doppler tracking of a spacecraft using its radio communications signal determines its orbit or trajectory from which the gravitational field and the planetocentric constant GM (where G is the universal gravitational constant and M is total mass) can be inferred. The Doppler-tracking method is highly accurate in the case of orbiting spacecraft, so that the nominal error of the deduced value of the mean density of the body, for example, 3933.5 ± 0.4 kg m^{-3} in the case of Mars, is primarily due to uncertainties in the value of G (Esposito et al., 1992). The mean density is indicative of the bulk composition of the terrestrial planets and terrestrial-type bodies. The density of a given material is strongly dependent on porosity and local pressure and temperature conditions, the latter of which may rise to a few hundreds of GPa and thousands of kelvins deep inside large planets like Venus and Earth. Uncompressed densities vary from about 1000 kg m^{-3} for planetary ices, to 3000 kg m^{-3} for silicate rocks, up to 8000 kg m^{-3} for metallic iron. The densities of the Moon, the Jovian satellite Io, and the inner terrestrial planets range from about 3300 to 5500 kg m^{-3}, suggesting that the Moon is predominantly composed of rocky material, whereas Mercury mostly consists of metal (Figure 2). The Mercury-sized large icy satellites Ganymede, Callisto, and Titan have densities of only about 2000 kg m^{-3}. This implies that ice and rock/metal are present in about equal amounts in their interiors. Enceladus, Triton,

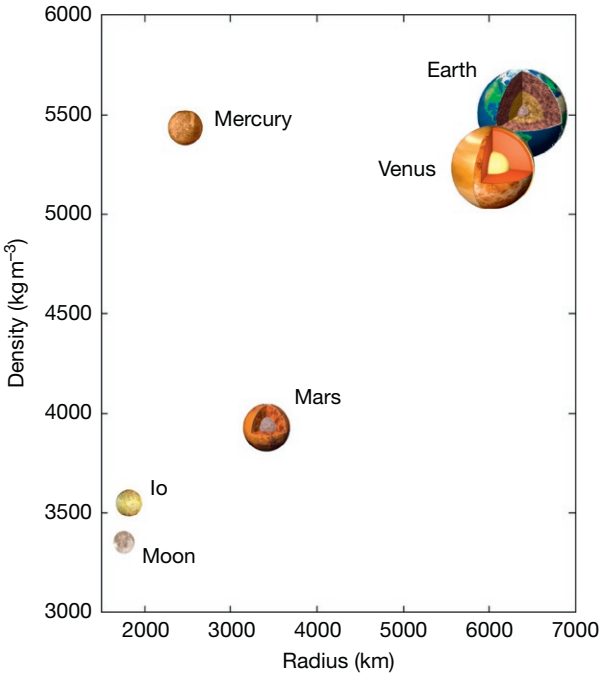


Figure 2 Radius–density relation of the terrestrial planets and the Moon. Note the unusually high mean density of Mercury implying that the planet's interior is predominantly composed of heavy elements such as iron.

Pluto, and Charon are icy bodies with similar ice–rock compositions (Hussmann et al., 2006).

The axial MoI of a planet or satellite C can be inferred from the precession of its rotation axis in the presence of external torques exerted on its equatorial bulge. Measurement of the precession rate might require the placement and long-term tracking of one or several landers on the surface of the body. For example, the combined analysis of Mars Global Surveyor (MGS) tracking and Mars Pathfinder and Viking lander range and Doppler data resulted in an improved value of the MoI factor of Mars, that is, the axial MoI normalized to the planet's mass times its radius squared, $C/MR_e^2 = 0.3644 \pm 0.0005$ (Konopliv et al., 2011), where R_e is Mars' equatorial radius fixed at 3396 km. This value is in good agreement with first determinations by Reasenberg (1977) and Kaula (1979) and much smaller than the MoI factor of 0.4 of a homogeneous sphere and is consistent with a significant concentration of mass toward the center due to a metallic core with a radius of about half that of the planet (Rivoldini et al., 2011; Schubert and Spohn, 1990; Sohl and Spohn, 1997; Sohl et al., 2005). The axial MoI of a planetary body can also be deduced from the figure of its gravitational field if the body is in hydrostatic equilibrium. However, this is a demanding requirement that is satisfied by only a few terrestrial-type bodies (see succeeding text).

10.02.2.2 Rotation and Tides

The internal mass distribution of a planetary interior can be inferred from its rotational and tidal response to external torques exerted on the planet or satellite if augmented by gravity field data (see **Chapter 10.04** in this volume). The torque applied to the body by the gravitational attraction of the Sun or a planet's satellite equals the inner product or tensor contraction of the body's inertia tensor and spin vector. Mercury and Venus have no satellites to affect their rotational states. The combined effect of the torques exerted by the tiny Martian moons, Phobos and Deimos, is much less than the torque exerted on Mars by the Sun (Van Hoolst et al., 2003). The torque applied to the Earth by the Moon decelerates the Earth's rotation due to tidal friction in the oceans and, to a minor extent, in the solid Earth (Platzman, 1984; Zharkov et al., 1996). In turn, the Moon is accelerated in its orbit and recedes from the Earth at a rate of a few centimeters per year. The rotation axes of Earth and Mars are tilted about axes normal to their orbital planes by 23.45°and 25.2°, respectively, thereby causing their spin axes to precess about the orbit normals under the influence of external gravitational torques.

Astronomical observations have revealed the precession rate of the Earth's spin axis and in turn the Earth's MoI. Radio tracking of the Viking and Mars Pathfinder landers, the stationary Opportunity Mars Exploration Rover, and MGS and Mars Odyssey orbiters has yielded similar information for Mars (Esposito et al., 1992; Folkner et al., 1997; Konopliv et al., 2006, 2011; Kuchynka et al., 2014; Yoder and Standish, 1997; Yoder et al., 2003). Laser ranging observations of the Moon, made possible by reflectors placed on the lunar surface during the Apollo program, have determined the Moon's rotational state (Dickey et al., 1994). Because of its orbital eccentricity and ellipsoidal shape, the Moon is subject to periodic changes in angular acceleration in response to gravitational torques exerted by the Sun and the Earth. The amplitude of the related 27-day forced libration in longitude has been used to infer the MoI of the Moon and the probable molten state of its central region (Khan et al., 2004; Williams et al., 2001; Yoder, 1981). In a similar way, measurements of Mercury's 88-day forced libration amplitude in longitude provide clues on the planet's internal mass distribution and coupling between core and mantle (Jehn et al., 2004; Margot et al., 2007, 2012; Peale, 1976, 1988). From a combined analysis of the rotational and low-degree gravitational field data, the radius of Mercury's outer liquid core and the presence of an inner solid core were inferred (Hauck et al., 2013; Rivoldini and Van Hoolst, 2013; Smith et al., 2012). It is unlikely, however, that the axial MoI of Venus could be derived from observations of its rotational state alone since the planet's retrograde rotation is extremely slow and its rotation axis is more or less perpendicularly aligned to the plane of its nearly circular orbit (Yoder, 1997).

10.02.2.3 Gravity and Topography

Gravimetric and magnetic methods utilize measurements of potential fields in the vicinity of planetary bodies. The long-wavelength part of the external gravitational field provides information on the structure of planetary interiors. Each element of mass in a planet or satellite contributes to the external gravitational field according to the Newton's law of gravitation. The gravitational force at a point exterior to the planet can be derived from the gradient or directional derivative of the gravitational potential caused by the total contribution of all mass elements when integrated over the entire volume of the body. In recognition of the first gravity experiments conducted by Galileo Galilei who measured fall times of objects possibly dropped from the leaning tower of Pisa, gravitational acceleration is given in units of $1\ \text{mGal} = 10^{-5}\ \text{m s}^{-2}$.

The gravitational fields of planetary bodies are measured from orbiting spacecraft. The technique uses the Doppler effect on radio signals transmitted from the Earth to the spacecraft and back. The frequency shift of the radio signals returned by the spacecraft relative to the signal emitted from the ground station is proportional to the velocity component along the direction of vision or line of sight and allows the calculation of the line-of-sight acceleration of the spacecraft (Dehant et al., 2009, 2011; Janle and Meissner, 1986). These data can be used to calculate the gravitational field at a resolution determined by the altitude and the frequency of signal transmission to the spacecraft. Laser altimetry from orbiting spacecraft has been used to determine the topography of the northern hemisphere of Mercury, the Moon, and Mars with high accuracy. Radar altimetry data obtained by the Magellan spacecraft have provided topographic maps of Venus. The correlation of topography with the distribution of the Bouguer gravity (gravity corrected for contributions of topographic masses above a well-defined reference level) can be used to infer the degree of isostatic compensation of topography and the depth at which stresses are fully compensated due to rheological and/or compositional changes. Gravity anomalies provide important constraints on elastic lithosphere thickness, crust thickness, crust density, and load density and can be interpreted in

terms of chemically or thermally induced lateral and depth variations of density. However, these interpretations suffer from the nonuniqueness inherent in any inversion of the gravity field (see **Chapter 10.05** in this volume).

The external gravitational field of a planet or satellite is generally described by a spherical harmonic expansion of the gravitational potential as a function of latitude, longitude, and radial distance to the center of mass. The frequency content of the amplitude spectrum of the gravitational potential gives insight into the rheology and the mechanical and thermal properties of the lithosphere, the outer rigid shell of a planet or satellite. The leading term in the spherical harmonic expansion of the external gravitational potential is the ratio between the planetocentric constant GM and the radial distance r representing the gravitational potential of the mass of the body concentrated in a point at its center of mass. The next terms of the spherical harmonic expansion are inversely proportional to r^3 and involve higher moments of the gravitational field caused by ellipsoidal distortions of the internal mass distribution. The magnitudes of these distortions are related to coefficients $J_2 = -C_{2,0}$ and $C_{2,2}$ for polar oblateness and equatorial ellipticity of the gravitational field, respectively. When the body is in hydrostatic equilibrium, its mean MoI can be deduced from the knowledge of these coefficients with important implications for the internal structure. However, if the lithosphere is thick and strong enough to support stresses associated with surface or subsurface mass anomalies, the planet or satellite may substantially deviate from hydrostatic equilibrium. Furthermore, mantle density anomalies and core–mantle boundary undulations may contribute to deviations from hydrostatic equilibrium. Accordingly, the Moon and Mercury are not expected to be fully hydrostatically compensated. The Earth is reasonably close to this state, but Mars is not, due to the uncompensated portion of the Tharsis uplift (Kaula, 1979; Reasenberg, 1977). Though the gravitational field of Venus is known quite well, it is not known if the planet is in hydrostatic equilibrium. Therefore, the gravitational field cannot be used to give a reliable value of the planet's MoI. A number of solid bodies in the outer solar system including the large icy satellites of Jupiter and Saturn, Ganymede, Callisto, and Titan, and the strongly tidally heated Jovian satellite Io are likely to be in hydrostatic equilibrium (see **Chapter 10.18** in this volume). The quadrupole gravitational fields of these Jovian satellites have been used to infer their moments of inertia (Schubert et al., 2004).

10.02.2.4 Magnetic Fields

Magnetic field observations provide constraints on the interior structure of planets and moons (see **Chapter 10.06** in this volume). The existence of an intrinsic planetary magnetic field on a global scale is a conclusive evidence that the body has a mobile, highly electrically conducting fluid region in its interior within which dynamo action creates the magnetic field. In the case of terrestrial planets, the site of magnetic field generation is a metallic core that is at least partially liquid. Earth and Mercury are the only terrestrial planets with a magnetic field that is unambiguously generated in its core. Mercury has a large iron core (Schubert et al., 1988) that is at least partially liquid (Margot et al., 2007) and a global magnetic field that must be generated therein (Anderson et al., 2011, 2012). Jupiter's icy satellite Ganymede is the only other terrestrial-like solar system body whose magnetic field is generated by an active liquid metallic core dynamo. The discovery of Ganymede's magnetic field provided crucial evidence that the satellite is differentiated and has a metallic core (Schubert et al., 1996). The absence of a planetary magnetic field, as in the cases of Venus and Mars, only implies the nonexistence of an active dynamo; it does not imply the nonexistence of a metallic core nor does it require a metallic core to be solid.

An external planetary magnetic field can be represented by a spherical harmonic expansion of its magnetic potential in a manner similar to the representation of the gravitational field. It is possible to identify the core dynamo field in the spectrum of the magnetic field spherical harmonic coefficients and thereby determine the radius of the core. This procedure has been used by Voorhies et al. (2002) to estimate the radius of the Earth's core as 3512 ± 64 km, in good agreement with the seismological radius (see also Elphic and Russell, 1978).

Though not in possession of global internally generated magnetic fields at present, the Moon (Halekas et al., 2001; Hood et al., 2001; Lin et al., 1998; Russell et al., 1973, 1975) and Mars (Acuña et al., 1998, 1999; Connerney et al., 2004) have localized magnetic fields associated with crustal remanent magnetization. On the assumption that the crustal magnetization was acquired when these bodies had active dynamos in the past, even the crustal magnetization is evidence for the existence of a metallic core in these bodies. This is particularly important for the Moon that has a small metallic core that is difficult to detect by any method (Garcia et al., 2011; Weber et al., 2011). Alternatively, Hood and Huang (1991) had proposed large-scale magnetization in lunar basin-forming impacts caused by plasma-induced antipodal amplification of ambient magnetic fields to explain the correlation of the largest magnetic field strengths with the antipodes of the largest lunar impact basins. The spatial variation of crustal magnetization and its associated magnetic fields also provide information on the internal structure of a body at shallow depth and the internal and surface processes that have affected the crust.

10.02.2.5 Electromagnetics

The phenomenon of electromagnetic induction can be utilized to probe planetary interiors. The method takes advantage of the electrically conducting nature of planetary materials and the time variability of the magnetic fields experienced by planetary bodies. If an electrically conducting object experiences a time-variable magnetic field, electrical currents are induced in the object (Faraday's law of induction). These induced currents in turn generate a magnetic field.

If the inducing and induced magnetic fields can be measured, then the electrical conductivity of the object can be inferred. This method has been successfully used to infer the electrical conductivity of the Earth (Banks, 1969; Chapman and Price, 1930; Hobbs, 1987; Kuvshinov and Olsen, 2006; Lahiri and Price, 1939; Rikitake, 1966; Roberts, 1986; Schuster, 1889; Tarits, 1994) and the Moon (Schubert and Schwartz, 1969; Sonett, 1982; Sonett et al., 1971). It has also been used to find subsurface liquid salt water oceans on the Galilean satellites Europa, Ganymede, and Callisto (Khurana et al.,

1998; Kivelson et al., 1999, 2000, 2002, 2004; Zimmer et al., 2000) and a subsurface magma ocean on Io (Khurana et al., 2011). In the case of the Earth, time variations in the external magnetic field are forced by variable ionospheric and magnetospheric current systems. The Moon experiences the variable magnetic field of the solar wind when it is outside the Earth's magnetosphere. The Galilean satellites experience the variability of the rotating, tilted Jovian magnetic field. On Earth, the external forcing magnetic field can be separated from the internal induced magnetic field by a network of surface magnetometers. More recently, satellite data have also been used for this purpose (Constable and Constable, 2004; Olsen, 1999). For the Moon, magnetometers on the surface and an orbiting spacecraft measured the solar wind forcing field and inductive response during the era of lunar exploration by Apollo spacecraft in the late 1960s and early 1970s. A magnetometer on the Galileo spacecraft measured the induced magnetic fields of Io, Europa, Ganymede, and Callisto. Knowledge of electrical conductivity inside a planetary body indirectly constrains the composition, temperature, and volatile content of the body through the dependence of the conductivity on these properties (Hood, 1986; Hood and Jones, 1987; Hood and Sonett, 1982; Hood and Zuber, 2000; Hood et al., 1982; Sonett et al., 1972). The electromagnetic sounding approach involves two inversion steps, inference of the electrical conductivity from magnetic measurements and inference of composition, temperature, and volatile content from the conductivity.

Magnetic field measurements obtained during the passage of the Moon through the Earth's geomagnetic tail provide a means of detecting the lunar core (Hood et al., 1999). For several days each month, the Moon passes through the north and south lobes of the geomagnetic tail and experiences a vacuum-like environment with a near steady uniform magnetic field. This results in an induced lunar magnetic dipole that acts to exclude the tail lobe field from the Moon's interior. The induced magnetic dipole is produced by electric currents that flow near the surface of the highly electrically conducting metallic core (Hood et al., 1999). The lunar-induced magnetic dipole has been detected by the magnetometer on the Lunar Prospector (LP) spacecraft; the measured value of the induced dipole moment implies a core of radius 340 ± 90 km (Hood et al., 1999) in good agreement with recent estimates based on reanalysis of Apollo seismic data (Garcia et al., 2011; Weber et al., 2011).

10.02.2.6 Seismology

The propagation of seismic waves following quakes or impacts can be used to infer the structure of planetary interiors. Lognonné (2005) and Lognonné and Mosser (1993) had provided reviews on planetary seismology (see **Chapter 10.03** in this volume). Whereas surface waves are confined to near-surface layers, body waves travel through the interior. The first body waves arriving at a seismic station are longitudinal or P-waves that involve material compression and rarefaction. The second body waves are transverse or S-waves involving shear motion perpendicular to the direction of propagation. Body waves are subject to reflection, transmission, and conversion from P- to S-waves and vice versa at internal boundaries at which prominent density and seismic velocity changes occur. A network of at least four seismic stations is required to infer the source location from a comparison of the seismic station records. Free oscillations on a global scale can be excited if the magnitudes of seismic events are sufficiently large.

For the Earth, seismology has not only revealed the existence of a metallic core surrounded by a rocky mantle but also shown that the core consists of a solid inner core composed of iron and nickel and a liquid outer shell containing a mixture of light elements. The analysis of surface waves has provided important clues on the chemical layering of the Earth's crust. Though seismic stations were placed on the Moon's surface during the Apollo exploration era, the collected data were unable at that time to confirm the presence or absence of a small metallic core and did not allow construction of detailed models of the lunar crust and mantle. As discussed later, a reanalysis of the Apollo seismic data has revealed more about the lunar interior. Seismic instruments placed on the surface of Mars by the Viking mission did not provide data that could be used to determine the planet's internal structure. A promising target for seismic exploration in the outer solar system is the Jovian satellite Europa owing to the presence of a satellite-wide internal liquid layer beneath its outer ice shell. A passive seismic experiment placed on the surface of Europa would be able to detect seismic activity generated by the formation of cracks, tidally induced quakes, or natural impacts. The thickness of the ice shell could be obtained from low-frequency band observations (0.1–10 Hz) of trapped surface waves (Kovach and Chyba, 2001; Lee et al., 2003; Panning et al., 2006).

10.02.2.7 Surface Geology and Composition

The geology and composition of a planet's surface provide important clues into what lies inside the body. These glimpses into the interior, though generally qualitative in nature, are valuable nonetheless and complement more quantitative constraints on internal structure.

The basaltic crust of Mercury and the basaltic lowlands of Venus are evidence of the internal differentiation of these bodies. The anorthositic lunar highlands not only attest to the differentiation of the Moon but also argue for a global-scale magma ocean early in lunar history (Smith et al., 1970; Warren, 1985; Wood et al., 1970). Mercury's high density suggests it is iron-rich, yet spectral evidence suggests that its surface is poor in oxidized iron (Blewett et al., 1997; Vilas, 1985), consistent with the segregation of its iron into a central core following differentiation under reducing conditions (Schubert et al., 1988; Sprague et al., 1994).

The crust dichotomy of Mars represents an ancient feature of early Noachian age (>3.5–3.7 Gy) that is preserved in the surface geology and tectonics, the cratering record, and the planet's gravity and magnetic field. The heavily cratered southern highland crust on Mars and the resurfaced northern lowlands of the planet indicate early differentiation and internal dynamic activity (Watters and McGovern, 2005), though removal of the northern crust by one or multiple giant impacts about 4 Gy has also been hypothesized (Wilhelms and Squyres, 1984). However, the detection of quasicircular depressions in the northern lowlands, interpreted as buried impact craters and basins, suggests a similar age for the lowland and the ancient

southern highland crust (Frey et al., 2002). This favors an endogenic origin of the crustal dichotomy involving dynamic processes like degree-1 mantle convection (Lingenfelter and Schubert, 1973; Roberts and Zhong, 2006; Schubert and Lingenfelter, 1973; Zhong and Zuber, 2001), thin-crust formation by early plate tectonics, and the solidification and convective instability of a pristine magma ocean on Mars (Elkins-Tanton et al., 2003b). The ancient dichotomy boundary was modified then by fluvial, aeolian, and glacial processes, as well as by widespread volcanic and sedimentary resurfacing of the northern lowlands (Watters and McGovern, 2005).

Surface geology and signs of endogenic activity are particularly important for inferring the internal structure of outer solar system moons. The heavily cratered surfaces of Jupiter's icy moon Callisto and Saturn's icy moon Rhea support the inference that these bodies are essentially undifferentiated (ice from rock) (Anderson and Schubert, 2007; Schubert et al., 2004). In contrast, the highly modified surfaces of the Jovian satellites Ganymede and Europa and the Saturnian satellite Enceladus are consistent with the separation of ice and rock in their interiors (Porco et al., 2006; Schubert et al., 2004, 2007). The ubiquitous and ongoing volcanic activity of Jupiter's satellite Io leaves little doubt that the moon has differentiated into an iron-rich core and silicate mantle (Moore et al., 2006). Thermal anomalies detected in Enceladus' south polar region and the active plumes spewing water from beneath the south polar surface of Enceladus (Porco et al., 2006; Spencer et al., 2006) provide additional persuasive evidence for the formation of a rock core and ice mantle inside this body.

While not definitive in themselves, surface geologic and compositional data help to constrain models of planetary and satellite interiors, especially when more quantitative observations are limited.

10.02.2.8 Material Properties

There are only a few kimberlite rock samples brought to the Earth's surface by violent volcanic eruptions that directly probe the deep interior of a terrestrial planet. Diamond deposits occasionally found in kimberlite rocks suggest that confining pressure must have exceeded the pressure at which diamonds are stable. Meteorites originally released from the surfaces of the Moon and Mars during one or several large impact events represent another important data source on the composition of terrestrial planet interiors. Much of our knowledge about the interior structure and evolution of the Earth and other terrestrial planets, however, comes from high-pressure mineral physics that includes laboratory experiments and computational studies. Bass (2004) summarized current and future research activities in this highly interdisciplinary and rapidly evolving field of research (see **Chapter 2.01**).

Earth and planetary material properties are affected by variable compositions and thermodynamic conditions. The core sulfur content may range from close, to eutectic, to iron-rich compositions with important implications for the physical state and density stratification of planetary cores (Balog et al., 2003; Fei et al., 1995, 1997; Kavner et al., 2001; Sanloup et al., 2000) due to the substantial melting point reduction with increasing sulfur concentration (Boehler, 1992, 1996a; Buono and Walker, 2011; Chudinovskikh and Boehler, 2007). Mantle rheology is affected by predominant mineral phase assemblages, water content, and typical mineral grain size distribution. It is possible that mantle mineralogies of other terrestrial bodies are not dominated by olivine but rather by mixtures of pyroxene minerals. Experimental and theoretical studies of two-phase rheologies (e.g., rock–volatile and rock–melt) are particularly useful for a better understanding of the influence of volatiles like water and carbon and solid gas hydrates on deeply situated geodynamic processes (Sloan, 2003), like the subduction of lithospheric plates on Earth, the rheological behavior of crust and mantle rocks, the kinetics of pressure-induced mineral phase transformations, and the extraction of partial mantle melts (Bass, 2004).

Nowadays, sophisticated tools are available to study the properties of matter under extreme pressure and temperature conditions that prevail in planetary interiors (Liebermann, 2005). From shock and static compression experiments using diamond-anvil facilities, densities, equation-of-state parameters (Hemley, 2006; Hemley and Ashcroft, 1998), and elastic properties of solid mineral phase assemblages (Anderson et al., 1992; Bina and Helffrich, 1992), as well as high-pressure properties of melts (Boehler, 1996b,c; Nishida et al., 2008), are derived. In these experiments, relatively small samples are subjected to conditions that may prevail deep in the Earth's mantle or even in its metallic core ($P > 135$ GPa, $T > 3000$ K). Future experiments are expected to provide important information on high-pressure transport properties like thermal conductivity and kinematic viscosity, the kinetics of chemical reactions at the core–mantle boundary, and mineral phase equilibriums up to several tens of GPa. While inelastic x-ray diffraction methods are used to measure sound velocities at very high pressures and temperatures, neutron scattering observations are well suited to determine the structural properties of silicate melts and aqueous solutions (Bass, 2004). Laboratory experiments need to be augmented by high-performance computations, for example, to simulate diffusion-limited transport processes occurring at atomic scales. These simulations are important to address the influence of oxygen fugacity on mantle rock phase transformation kinetics and the element solubility of light-element admixtures in liquid iron alloys (Badding et al., 1991; Wood, 1993).

10.02.3 Interior Structure and Composition

In the following, we consider two- and three-layer structural models before we discuss radially symmetrical, depth-dependent interior structure models.

10.02.3.1 Two- and Three-layer Structural Models

Models of the internal density distribution of terrestrial planets suffer from an inherent nonuniqueness since there are usually fewer constraints than unknowns. These models are required to satisfy two constraints, the mean density ρ as derived from the total radius R and mass M and the mean MoI I that can be determined from the quadrupole moments of the gravitational field and the precession rate of the spin axis.

The mean density of a two-layer spherical body is given by

$$\rho R^3 = (\rho_c - \rho_m)R_c^3 + \rho_m R^3 \quad [1]$$

from which the relative core radius R_c/R, core mass fraction M_c/M, and dimensionless mean MoI factor are obtained according to

$$\frac{R_c}{R} = \left(\frac{\rho - \rho_m}{\rho_c - \rho_m}\right)^{1/3} \quad [2]$$

$$\frac{M_c}{M} = \frac{\rho_c}{\rho}\left(\frac{R_c}{R}\right)^3 = \frac{\rho_c}{\rho}\frac{(\rho - \rho_m)}{(\rho_c - \rho_m)} \quad [3]$$

and

$$\text{MoI} = \frac{I}{MR^2} = \frac{2}{5}\left(\frac{(\rho - \rho_m)^{5/3}}{\rho(\rho_c - \rho_m)^{2/3}} + \frac{\rho_m}{\rho}\right) \quad [4]$$

respectively. For stably stratified layering, the corresponding increase of density with depth is equivalent to the requirement $\rho_c > \rho$ and $0 < \rho_m < \rho$. In Figure 3, contours of MoI are shown for plausible ranges of core and mantle densities normalized to the mean density, ρ_c/ρ and ρ_m/ρ, respectively.

For a rapidly rotating planetary body in hydrostatic equilibrium, the mean MoI as a measure for mass concentration toward the center can be obtained from the dimensionless polar MoI factor, C/MR^2, and the second-degree coefficient of the spherical harmonic expansion of the gravity field, J_2, by

$$\frac{I}{MR^2} = \frac{C}{MR^2} - \frac{2}{3}J_2 \quad [5]$$

where

$$J_2 = \frac{1}{MR^2}\left[C - \frac{A+B}{2}\right] = -C_{2,0} \quad [6]$$

is the gravitational oblateness and A, B, and C are the planet's principal equatorial and polar moments of inertia, where $A < B < C$, respectively (Kaula, 1979; Reasenberg, 1977). A more comprehensive account including minor nonhydrostatic corrections to the mean MoI factor due to symmetrical distributions of topographical masses, like the Tharsis rise on Mars, is provided by Sohl et al. (2005).

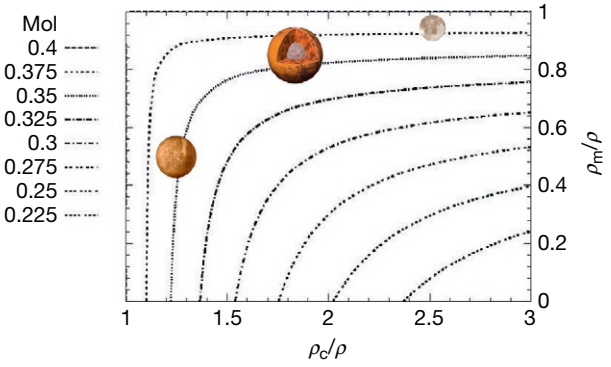


Figure 3 Contours of the mean moment of inertia factor MoI for two-layer structural models of planetary interiors as a function of core and mantle density ρ_c and ρ_m, respectively, relative to mean density ρ. The symbols shown in the diagram represent putative locations of the Moon, Mars, and Mercury.

The gravitational acceleration g_r and hydrostatic pressure p_r as a function of radial distance from the center of the planet r are given by

$$g_r = \frac{4}{3}\pi G r \rho_c \quad 0 \le r \le R_c$$

$$= \frac{4}{3}\pi G r\left[\rho_m + (\rho_c - \rho_m)\left(\frac{R_c}{r}\right)^3\right] \quad R_c \le r \le R \quad [7]$$

and

$$p_r = \frac{4}{3}\pi G \rho_m R_c^3(\rho_c - \rho_m)\left[\frac{1}{r} - \frac{1}{R}\right] + \frac{2}{3}\pi G \rho_m^2(R^2 - r^2) \quad R_c \le r \le R$$

$$= \frac{2}{3}\pi G \rho_c^2(R_c^2 - r^2) + \frac{2}{3}\pi G \rho_m^2(R^2 - R_c^2) + \frac{4}{3}\pi G \rho_m R_c^3(\rho_c - \rho_m)\left[\frac{1}{R_c} - \frac{1}{R}\right] \quad 0 \le r \le R_c \quad [8]$$

respectively, where G is the gravitational constant (Turcotte and Schubert, 2002).

Given the paucity of information available, three-layer models representing the core, the mantle, and the crust can be considered as a useful approximation to the structure of terrestrial planet interiors. The corresponding structural equations are given by

$$\rho = \rho_s + (\rho_c - \rho_m)\left(\frac{R_c}{R}\right)^3 + (\rho_m - \rho_s)\left(\frac{R_m}{R}\right)^3 \quad [9]$$

$$\text{MoI} = \frac{I}{MR^2} = \frac{2}{5}\left(\frac{\rho_s}{\rho} + \frac{\rho_c - \rho_m}{\rho}\left(\frac{R_c}{R}\right)^5 + \frac{\rho_m - \rho_s}{\rho}\left(\frac{R_m}{R}\right)^5\right) \quad [10]$$

where the core radius R_c, the crust–mantle radius R_m, the crust density ρ_s, the mantle density ρ_m, and the core density ρ_c are unknown. Even two-layer interior structure models lacking a crust layer ($\rho_s = 0$, $R_m = R$) would have fewer constraints than unknowns, that is, ρ_c, R_c, and ρ_m. We refer the reader to Sohl et al. (2009) for a compilation of two- and three-layer structural model determinations of the terrestrial planets, the Moon, and selected Jovian and Saturnian satellites in the tabular form.

10.02.3.2 Depth-Dependent Structural Models

The equation for the determination of the density distribution in the regions of the Earth with adiabatic temperature profiles was obtained by Williamson and Adams (1923). The first current models of the Earth were constructed at the beginning of the 1940s by Bullen (1947), who introduced the subdivision of the Earth's interiors into characteristic zones that is still valid today. Deviations from adiabaticity, the presence of phase transition zones, and changes of the chemical composition result in additional terms to the Williamson–Adams equation. These effects are described in detail for the Earth's interior in the first chapter of Zharkov et al. (1996).

A commonly used approach to modeling the interior structure of a terrestrial planet given its mass, radius, and MoI (see Wood et al. (1981) for a review) converts the estimated bulk composition into a mineralogical model and adopts potential temperatures for

the mantle layers and the core from thermal history calculations. The potential temperature is the mantle temperature extrapolated adiabatically to the surface pressure. The radial density distribution is then calculated under the assumption of hydrostatic and thermal equilibrium using an equation of state to correct for compression and thermal expansion. The total mass and the MoI are calculated from the model and compared with the data. This approach can be modified by simultaneously calculating the thermal and mechanical structure of a terrestrial body (e.g., Sohl and Spohn, 1997; Wagner et al., 2011, 2012). Standard-temperature-and-pressure values of the density, the bulk modulus, and the thermal expansivity are calculated from laboratory data for the individual layers of the model.

10.02.3.2.1 Governing equations

If a spherically symmetrical planet in perfect mechanical equilibrium and thermal steady state is assumed, the following set of differential equations for mass m, iron mass m_{Fe}, mean MoI I, acceleration of gravity g, pressure p, and heat flux q can be derived from fundamental principles (Wagner, 2014):

$$\frac{\mathrm{d}m}{\mathrm{d}r} = 4\pi r^2 \rho_{\mathrm{r}} \quad [11]$$

$$\frac{\mathrm{d}m_{\mathrm{Fe}}}{\mathrm{d}r} = x_{\mathrm{Fe}} \frac{\mathrm{d}m}{\mathrm{d}r} \quad [12]$$

$$\frac{\mathrm{d}I}{\mathrm{d}r} = \frac{8}{3}\pi r^4 \rho_{\mathrm{r}} \quad [13]$$

$$\frac{\mathrm{d}g}{\mathrm{d}r} = 4\pi G \rho_{\mathrm{r}} - 2\frac{g}{r} \quad [14]$$

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho_{\mathrm{r}} g \quad [15]$$

$$\frac{\mathrm{d}q}{\mathrm{d}r} = \rho_{\mathrm{r}} \varepsilon_{\mathrm{r}} - 2\frac{q}{r} \quad [16]$$

where r is the radial distance from the center of the planet, G is the gravitational constant, ρ is the density, x_{Fe} is the concentration of iron per unit mass, and ε is the specific heat production rate. The subscript r indicates quantities that are local functions of p, T, and composition.

Heat is primarily carried by conduction across the stagnant outer portion of a planet's silicate shell and the top and bottom thermal boundary layers of mantle convection. The corresponding radial temperature gradient is given by

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{q}{k_{\mathrm{r}}} \quad [17]$$

where k is the thermal conductivity. Within the convective portion of the silicate shell and the outer liquid core shell, the temperature gradient can be approximated by the adiabatic temperature gradient (Stacey, 1977)

$$\frac{\mathrm{d}T}{\mathrm{d}r} = T\frac{\gamma_{\mathrm{r}}}{K_{S,\mathrm{r}}}\frac{\mathrm{d}p}{\mathrm{d}r} = T\frac{\gamma_{\mathrm{r}}\rho_{\mathrm{r}}}{\Phi_{\mathrm{r}}}\frac{\mathrm{d}p}{\mathrm{d}r} = \left(\frac{\mathrm{d}T}{\mathrm{d}r}\right)_{\mathrm{ad}} \quad [18]$$

where $\gamma = \alpha K_S/\rho c_P$ is the thermodynamic Grüneisen parameter, α is the thermal expansivity, c_P is the specific heat, and K_S is the adiabatic bulk modulus defined as

$$\frac{1}{K_S} = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial p}\right)_S \quad [19]$$

Finally, p, K_S, and the shear modulus G_μ are related to the seismic velocities

$$V_{P,\mathrm{r}} = \sqrt{\frac{K_{S,\mathrm{r}} + \frac{4}{3}G_{\mu,\mathrm{r}}}{\rho_r}} \quad [20]$$

and

$$V_{S,\mathrm{r}} = \sqrt{\frac{G_{\mu,\mathrm{r}}}{\rho_r}} \quad [21]$$

of longitudinal P- and transversal S-waves, respectively, from which the seismic parameter

$$\Phi_{\mathrm{r}} = \frac{K_{S,\mathrm{r}}}{\rho_{\mathrm{r}}} = V_{P,\mathrm{r}}^2 - \frac{4}{3}V_{S,\mathrm{r}}^2 \quad [22]$$

can be derived.

The radial positions of pressure-induced mantle phase boundaries like the exothermic olivine–β-spinel (wadsleyite) and β-spinel (wadsleyite)–γ-spinel (ringwoodite) transitions and the endothermic γ-spinel (ringwoodite)–perovskite transition can be obtained from the intersections of the temperature profile calculated from eqns [16] to [18] and the specific Clausius–Clapeyron curves.

The set of basic differential eqns [11]–[18] can be separated into two subsets that are coupled through the density ρ. The mechanical properties of the interior are calculated from eqns [11] to [15], while eqns [16]–[18] give the thermal structure of the model. These equations have to be supplemented with an appropriate equation of state to include pressure-induced compression and thermal expansion effects on the density and are discussed in more detail in the succeeding text (Stacey et al., 1981) and illustrated in **Figure 4**.

10.02.3.2.2 Boundary conditions

The set of basic differential eqns [11]–[18] can be solved by numerical integration with respect to the following boundary conditions. The central boundary conditions at $r = 0$ are

$$m = 0;\ m_{\mathrm{Fe}} = 0;\ \theta = 0;\ g = 0;\ p = p_{\mathrm{c}};\ q = 0;\ T = T_{\mathrm{c}} \quad [23]$$

The surface boundary conditions at $r = R$ are

$$m = M;\ m_{\mathrm{Fe}} = M_{\mathrm{Fe}};\ \theta = I;\ g = g_{\mathrm{p}};\ p = p_{\mathrm{p}};\ q = q_{\mathrm{p}};\ T = T_{\mathrm{p}} \quad [24]$$

While the mass M and the mean surface values of gravity g_{p}, pressure p_{p}, and temperature T_{p} are derived from spacecraft and Earth-based observations, the surface heat flux q_{p} of most terrestrial bodies other than the Earth is unknown at present. Since there are three observational constraints on the model, the mass M, the radius R, and the mean MoI I, respectively, three parameters are adjustable, the values of which are iteratively adjusted such that the observational constraints can be satisfied. These parameters are the central pressure p_c, the central temperature T_c, and the pressure at the core–mantle boundary p_{cmb}. For the sake of rapid convergence to the successful solution, the initial set of starting parameters is constrained to lie sufficiently close to the final set of parameters. An educated guess of the initial set can be obtained by using the previously mentioned analytic solutions for a three-layer structural model, having the planet's mass, iron mass fraction,

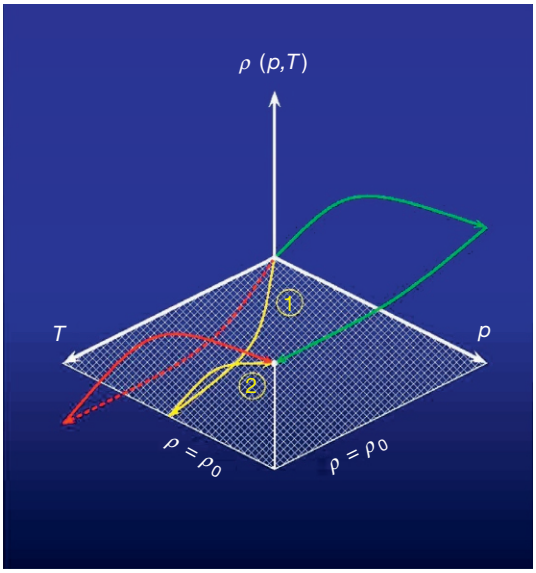


Figure 4 Schematic view of the extrapolation of density $\rho(p, T)$ to elevated pressures p and temperatures T. The origin of the coordinate system refers to an arbitrarily chosen reference density $\rho_0(p_0, T_0)$ at fixed (standard) pressure and temperature conditions p_0 and T_0, respectively. First, the increase of thermal pressure caused by thermal expansion is calculated (path 1); then, the isothermal Birch–Murnaghan equation (path 2) is applied to account for pressure-induced compression. Also shown are alternate thermodynamic paths. Modified from Wood JA, Anderson DL, Buck WR, et al. (1981) Geophysical and cosmochemical constraints on properties of mantles of the terrestrial planets. In: Project BVS (ed.) *Basaltic Volcanism on the Terrestrial Planets*, pp. 633–699. New York: Pergamon.

basaltic crust density, and homogeneously distributed crust/mantle heat sources (Turcotte and Schubert, 2002).

10.02.3.2.3 Equation of state

A convenient method to calculate density as a function of depth, first applied to seismological data from the Earth's deep interior, relies on the assumption of hydrostatic pressure (eqn [15]) and adiabatic temperature (eqn [18]) conditions and is sometimes referred to as the *Williamson–Adams* method honoring its authors. The resultant adiabatic density gradient

$$\frac{\mathrm{d}\rho}{\mathrm{d}r} = -\frac{\rho_\mathrm{r} g_\mathrm{r}}{\Phi_\mathrm{r}} = \Phi_\mathrm{r}^{-1}\frac{\mathrm{d}p}{\mathrm{d}r} \qquad [25]$$

can be readily obtained from insertion of eqns [19] and [22] in [15] and rearranging terms. Later, a correction term was added to address temperature deviations from an adiabatic reference state. These are caused by thermal boundary layers of mantle convection and/or compositional changes and pressure-induced mineral phase transformations in a silicate mantle, giving rise to superadiabatic temperature gradients

$$\tau = \frac{\mathrm{d}T}{\mathrm{d}r} - \left(\frac{\mathrm{d}T}{\mathrm{d}r}\right)_\mathrm{ad} \qquad [26]$$

that may profoundly affect the density stratification. The overall density gradient is then given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}r} = \Phi_\mathrm{r}^{-1}\frac{\mathrm{d}p}{\mathrm{d}r} + \alpha_\mathrm{r}\rho_\mathrm{r}\tau_\mathrm{r} \qquad [27]$$

and can be derived from the ambient density ρ_r and the seismic parameter Φ_r if adiabatic compression dominates (first term). In the event of perfect adiabaticity, τ_r will equal zero, and the superadiabatic correction (second term) would vanish.

To interpret density profiles inferred from seismological observations in terms of compositional changes and temperature anomalies, it is necessary to consider decompressed values of density. Reliable extrapolation to zero-pressure and room-temperature conditions requires laboratory measurements of thermodynamic material properties be experimentally determined at high pressures and temperatures. Extrapolation between different thermodynamic states represents a principal source of uncertainty in density modeling because of incomplete knowledge of the material parameters involved. For technical reasons, laboratory measurements are usually carried through under isothermal conditions.

The isothermal bulk modulus is then defined as

$$\frac{1}{K_T} = \frac{1}{\rho}\left(\frac{\partial\rho}{\partial p}\right)_T \qquad [28]$$

and can be related to K_S by using the isothermal–adiabatic transformation

$$\frac{K_S}{K_T} = 1 + \gamma\alpha T \qquad [29]$$

if the Grüneisen parameter γ and thermal expansivity α are specified. A linearly pressure-dependent expression for K_T is frequently used, that is,

$$K_T = K_{0T} + \frac{\mathrm{d}K_{0T}}{\mathrm{d}p}p = K_{0T} + K_0' p \qquad [30]$$

and taken as $\gamma = \gamma_0(\rho/\rho_0)^{-l}$, with index 0 referring to an arbitrarily chosen reference state at fixed (standard) pressure and temperature conditions p_0 and T_0, respectively, l = const., and constant first pressure derivative K_0'. Insertion of eqn [30] in [28] and subsequent integration yield

$$p = \frac{K_{0T}}{K_0'}\left[\left(\frac{\rho}{\rho_0}\right)^{K_0'} - 1\right] \qquad [31]$$

which is occasionally referred to as *Murnaghan's* equation. The corresponding density relative to its reference value is then given by

$$\frac{\rho}{\rho_0} = \left(1 + \frac{K_0'}{K_{0T}}p\right)^{1/K_0'} \qquad [32]$$

The density distribution within a chemically homogeneous layer was extended by Birch (1952) who suggested use of a higher-order isothermal finite-strain equation of state to correct for pressure-induced compression and to apply temperature corrections through calculation of the thermal pressure contribution (Stacey et al., 1981).

For example, an isothermal *Birch–Murnaghan* equation truncated at third-order in strain involves K_{0T} and K_0' to correct for pressure-induced compression and the product $\alpha_0 \times K_{0T}$ to correct for thermal pressure effects, according to (Anderson and Baumgardner, 1980)

$$p = \frac{3K_{0T}}{2}\left[\left(\frac{\rho}{\rho_0}\right)^{7/3} - \left(\frac{\rho}{\rho_0}\right)^{5/3}\right] \left\{1 + \frac{3}{4}\left(K_0' - 4\right)\left[\left(\frac{\rho}{\rho_0}\right)^{2/3} - 1\right]\right\} + p_{\text{th}} \quad [33]$$

Since internal temperatures of the terrestrial planets usually exceed the Debye temperature, the quasiharmonic approximation is valid (Anderson et al., 1992) and the thermal pressure given by

$$p_{\text{th}} = \int_{T_0}^{T}\left(\frac{\partial p}{\partial T}\right)_V \mathrm{d}T = \int_{T_0}^{T} \alpha K_T \mathrm{d}T \approx \alpha_0 K_{0T}(T - T_0) \quad [34]$$

can be used in eqn [33] to include the temperature effect in the equation-of-state parameters (Anderson, 1984). We refer the reader to Sohl et al. (2009) for a compilation of equation-of-state parameters of selected Mg- and Fe-rich mantle mineralogical phase assemblages and core constituents and references therein.

In summary, the simplicity of the *Williamson–Adams* method is partly counterbalanced by the disadvantage that seismological data are not frequently available for terrestrial planets other than the Earth and the Moon. The linear pressure dependence of K_T inherent in Murnaghan's equation may represent an adequate assumption for small bodies like the Moon and the satellites of the outer giant planets (see **Chapter 10.18** in this volume). However, even for planets in the mass and size range of Mercury or Mars, with central pressures of the order of the isothermal bulk modulus, more reliable higher-order finite-strain equations of state corrected for thermal pressure effects such as eqn [33] are recommended for the construction of models of the internal density distribution.

10.02.4 Earth as a Type Example of a Terrestrial Planet

Inferences about the structure, composition, and mineralogy of the planets are guided by our knowledge of the Earth. In some cases, Venus, for example, observational constraints on the interior are so minimal that the best we can do is to argue by analogy with Earth. Our knowledge of the Earth's interior, though partially based on the same types of observations available for the planets, rests largely on seismological data. The Moon is the only planetary body other than Earth for which we have seismological data, and those data are limited in both quality and quantity. The structure, composition, and mineralogy of the Earth are discussed at length in other volumes of this treatise (see **Chapter 1.08** and **Chapter 2.01**) and in *Treatise on Geochemistry*. The brief summary we present here is intended only to place the blurry pictures of planetary interiors in the context of the sharper picture of the Earth's interior.

10.02.4.1 General

The basic structure of the Earth is that of a three-layer sphere. The innermost layer is a metallic core of radius 3486 km. Surrounding the core is a rocky spherical shell or mantle of thickness 2885 km, and encircling the mantle is a thin rocky crust distinct in composition from the underlying mantle. The core is itself divided into two parts, an inner solid core of radius 1217 km surrounded by an outer liquid shell about 2269 km thick. It is believed that the inner core has formed as a consequence of the cooling of the Earth over geologic time, a process that has resulted in the partial solidification of the core from the inside out. The mantle is itself divided into two parts, the upper mantle about 660 km thick and the lower mantle about 2225 km thick. The mantle subdivisions are based on the occurrence of solid–solid phase transitions in the mantle rock, about which more is said in the succeeding text. The crust is uniform in neither composition nor thickness.

An ~6 km thick crustal layer covers the floors of the oceans, while a compositionally distinct crust of about 30 km thickness comprises the continents. The basic structure of the Earth is shown in **Figure 5**.

10.02.4.2 Interior Structure

The overall structure of the Earth is generally believed to have been set early in Earth's evolution, within tens to about 100 My after accretion was complete (see Volume 9, 'Evolution of the Earth'). The gravitational potential energy released upon accretion was large enough to melt the Earth's interior yielding liquid metal, mostly iron, that sank toward the center and accumulated to form the metallic core. The process of core formation could have begun before accretion was complete. Core formation itself releases additional energy that could have contributed to melting the mantle. Formation of the core could have been a runaway process. The timing of crustal formation is more uncertain. Certainly, oceanic crust (basalt) is being produced today by melting of mantle rocks at mid-ocean ridges. This process has occurred throughout Earth's history, as long as plate tectonics has been active. Basaltic volcanism does not require plate tectonics, as is evident from the surfaces

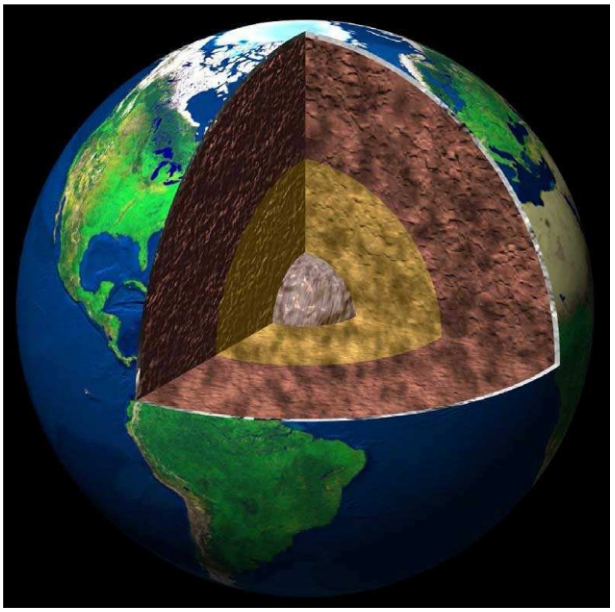


Figure 5 Cutaway view of the Earth's interior. © Calvin J. Hamilton.

of other planets, so basaltic crust was likely produced on Earth early in its evolution by plate tectonics or other processes. Oceanic crust is recycled back into the mantle by plate tectonics. The oldest oceanic crust on Earth is only about 200 My. The more difficult problem in crustal evolution is the history of the continental crust. The more siliceous continental crust is believed to have been produced by multiple stages of melting in the presence of water. The process is occurring in the andesitic volcanoes at convergent plate margins. The main question is how much continental crust existed on Earth as a function of time throughout Earth's history.

10.02.4.3 Composition

The bulk composition of the Earth, aside from its volatiles, is believed to be that of the meteorites classified as CI chondrites, the assumed primordial building blocks of the Earth. Accordingly, the Earth is 32.7 wt.% Fe, 15.4 wt.% Mg, 14.2 wt.% Si, 1.89 wt.% Ni, 1.71 wt.% Ca, and 1.59 wt.% Al (McDonough and Sun, 1995). Other elements are of course present in relatively smaller amounts. Much of the Earth's Fe and Ni have segregated into the core that contains about 87.5 wt.% Fe, 5.4 wt.% Ni, 0.95 wt.% Cr, 0.5 wt.% Mn, and unknown amounts of light elements (McDonough and Sun, 1995). In descending order by their significance, these elements are Si, S, O, H, and C (Poirier, 1994) (see **Chapter 2.06** in Volume 2 of this treatise). The light elements of the core are preferentially excluded from the inner core upon its solidification and concentrated in the liquid outer core. The major elements of the mantle rocks, aside from volatiles, are Mg (22.8 wt.%), Si (21.0 wt.%), Fe (6.26 wt.%), Ca (2.53 wt.%), and Al (2.35 wt.%) (McDonough and Sun, 1995).

The composition of the mantle is also constrained by its density, seismic velocities, seismic anisotropy, and the compositions of ophiolite complexes and mantle xenoliths brought to the surface by kimberlitic and alkali basaltic eruptions. The elements of the mantle are contained in the minerals olivine ($(Mg,Fe)_2SiO_4$), orthopyroxene ($(Mg,Fe)SiO_3$), clinopyroxene ($([Ca,Mg]_2,NaAl)Si_2O_6$), and garnet ($(Mg,Fe,Ca)_3Al_2Si_3O_{12}$). Olivine is a complete solid solution of Mg and Fe silicates with end-members fayalite (Fe_2SiO_4) and forsterite (Mg_2SiO_4). Orthopyroxene is a limited solid solution with magnesium end-member enstatite (the Fe end-member is unstable). Clinopyroxene is a pyroxene solid solution with Ca and Al. Among all possible rock assemblages of these minerals, only peridotites (olivine + pyroxene) and eclogites (pyroxene + garnet) are commonly found in mantle-derived samples. Eclogite is isochemical with basalt and transforms to basalt at depths of less than about 80 km in the Earth.

The mineralogy and composition of the Earth's mantle might be that of pyrolite, a peridotite model introduced in 1962 by Ringwood to explain the seismic, petrologic, and mineralogical properties of the upper mantle (Ringwood, 1975). To better understand the structure of the Earth's mantle, the olivine phase diagram obtained from experiments was of great importance (Akimoto and Fudisawa, 1968; Ringwood, 1970). This diagram allowed explanation of the boundaries between different layers in the mantle and provided constraints on the radial temperature distribution.

10.02.4.4 Mineralogy

Olivine and pyroxene transform to higher density polymorphs under the high temperatures and pressures encountered at depth in the Earth's mantle. At a depth of about 410 km, olivine transforms to spinel (the transformation occurs in two steps involving wadsleyite, or β-spinel, and at higher pressure ringwoodite, or γ-spinel). A prominent seismic discontinuity at the depth of about 410 km is believed associated with the olivine–spinel phase change. This subsolidus phase change is exothermic and involves jumps in seismic velocities and density. Another major seismic discontinuity occurs at the depth of about 660 km, and this is believed to coincide with the transformation of ringwoodite to magnesium perovskite and magnesiowüstite. This solid–solid phase transformation is endothermic. The region between 410 and 660 km depth is known as the transition zone. The depth of 660 km marks the boundary between the upper mantle and the lower mantle. Perovskite-forming reactions also occur in the pyroxene system but over a wider pressure interval than in the olivine system. Garnet, for example, dissociates to form $MgSiO_3$ and $CaMgSiO_3$ perovskites plus Al_2O_3. The lower mantle is thus dominated by silicate perovskites.

Recent laboratory and theoretical results have demonstrated that perovskite transforms to still another structure, postperovskite, at pressures and temperatures found just above the core–mantle boundary in the Earth (Murakami et al., 2004; Oganov and Ono, 2004; Tsuchiya et al., 2004). The perovskite–postperovskite transformation, like the olivine–spinel phase change, is an exothermic reaction. The location of this phase transition is coincident with the several hundred kilometer thick layer at the bottom of the mantle known as the D" layer wherein seismic velocities undergo large variations (Lay et al., 2005).

The subsolidus phase changes in the Earth's mantle have important effects on the dynamics of the mantle (Schubert et al., 2001). The exothermic reactions generally promote convection, while the endothermic phase change retards it. Do similar phase changes occur in other terrestrial planets? Venus is large enough that the olivine–spinel and spinel–perovskite phase changes should be present in its mantle. However, the pressure at the base of the Venusian mantle may not be sufficiently high for the occurrence of the perovskite–postperovskite phase change. Mars is so small that only the olivine–spinel phase change might occur near the base of its mantle. Possible phase changes in the mantles of the other planets will be discussed in more detail in the succeeding text.

10.02.5 Vesta

NASA's Dawn spacecraft orbited the asteroid 4 Vesta for about a year returning valuable information about its surface and interior. Vesta, formally known as 4 Vesta, is the second-largest asteroid in the solar system (**Figure 6**). Vesta's shape is approximated by a triaxial ellipsoid with dimensions 286.3, 278.6, and 223.2 km; its mean radius is 262.7 km (Russell et al., 2012). Prior to the arrival of the Dawn spacecraft at Vesta, the asteroid was considered to be a differentiated protoplanet with

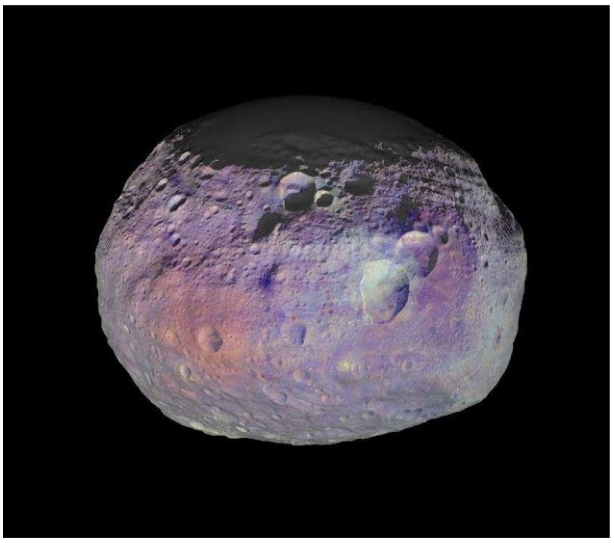


Figure 6 Image of Vesta captured by the Dawn spacecraft (PIA15506). The surface mineralogy of Vesta (colorized) indicates a complex magmatic evolution in the course of internal differentiation and crust formation. © NASA/JPL–Caltech/UCLA/MPS/DLR/IDA/PSI.

an iron core that survived essentially intact from the earliest stages of solar system formation (Kleine et al., 2009; Righter and Drake, 1997). This view was based on the identification of 4 Vesta as the source of the howardite–eucrite–diogenite (HED) meteorites. All the data acquired by Dawn are consistent with this paradigm.

The mass and average density of Vesta are 2.59×10^{20} kg and 3456 kg m^3, respectively (Russell et al., 2012). The average density of Vesta, combined with its measured second-degree gravitational moment $J_2 = 0.03178$ (Russell et al., 2012) and reasonable assumptions about iron core density in a two-layer model, leads to values of core radius and core mass fraction of about 110 km and 18%, respectively, consistent with estimates from analyses of HED meteorites (Russell et al., 2012). Based on its shape, size, mean density, and rotation rate ($1617.333^\circ\,day^{-1}$), it is unlikely that Vesta is in hydrostatic equilibrium.

The surface mineralogy of Vesta is similar to the composition of the HED meteorites, confirming that Vesta's crust formed by melting of a chondritic parent body. Vesta is scarred by a giant impact basin, Rheasilvia, at its south pole. The basin is about 500 km in diameter and about 19 km deep. It has a large central peak almost as high as Olympus Mons on Mars. Rheasilvia obliterated about half of an older 400 km south polar basin, Veneneia (Schenk et al., 2012). Both basins are geologically young, about 1–2 billion years. The volume of material excavated by these impacts is sufficiently large to produce the Vesta-family asteroids (vestoids or asteroids with orbits and reflectance spectra similar to Vesta) and the HED meteorites.

These impacts produced a strong dichotomy between the northern and the southern hemispheres of Vesta, reflected in surface albedo and crater densities. Vesta's early differentiation of an iron core was likely caused by the heating of the short-lived radioactive isotopes ^{26}Al and ^{60}Fe. The half-lives of these rapidly decaying elements place a tight constraint on the timing of the formation of Vesta. The discovery of remanent magnetization in the Millbillillie eucrite and ALH 81001 (Fu et al., 2012) suggests the existence of an early dynamo on their parent body 4 Vesta.

10.02.6 The Moon

The knowledge of the lunar interior is one of the key elements to a better understanding of the formation and evolution of the solar system. The earliest history of the Moon together with the first billion years of the impact history of the Earth–Moon system is partly retained in the subsurface structure and surface geologic record of the Moon. Dynamic modeling and isotopic data suggest that the accretion and differentiation of the Moon occurred soon after the formation of a vapor and debris cloud caused by the giant impact of a Mars-sized planetesimal into the early Earth (Canup and Asphaug, 2001). The chronology of major impact events has been derived from rock and soil samples collected during the Apollo and Luna missions and represents the current best database for dating other planetary surfaces in the solar system (Kaula et al., 1986). A number of lunar meteorites provide additional constraints on the early evolution and bulk composition of the lunar crust (Korotev, 2005).

10.02.6.1 General

The low mean lunar density of 3344 kg m^{-3} suggests that the Moon is depleted in iron relative to the Earth and any other terrestrial planet. Despite seismic measurements at the Apollos 12, 14, 15, and 16 landing sites and improved values of the MoI and tidal potential Love number, models of the lunar interior are nonunique. Some uncertainty is still connected with the radius and physical state of a small iron core, the possible existence of an inner solid core, more gradual or discontinuous seismic velocity variations within the lunar mantle, and lateral and vertical heterogeneities of the lunar crust. There is a pronounced dichotomy between the crustal thicknesses of the lunar nearside and farside as a consequence of the early differentiation of the Moon, possibly augmented by variable heat transfer in the molten lunar interior and large-scale insulation due to the formation of a protocrust and subsequent deposition of thick impact ejecta (Jolliff et al., 2000b). The internal differentiation of the Moon was accompanied by the formation of a magma ocean subsequent to hot accretion and extraction of a highly aluminous flotation crust enriched in plagioclase feldspar (Smith et al., 1970; Warren, 1985; Wood et al., 1970). The duration and depth of differentiation during the magma-ocean phase, the possible existence of an undifferentiated lower mantle, and the mechanism of core formation are among the key questions of lunar science (Jolliff et al., 2000b). For review articles on the geophysics, geochemistry, and geology of the Moon, we refer the reader to Wieczorek et al. (2006), Shearer et al. (2006), and Jaumann et al. (2012), respectively.

10.02.6.2 Interior Structure

The Clementine and LP missions have returned the first global data sets of lunar gravity, topography, remanent magnetism,

mineralogy, and chemical composition of the lunar surface (Binder, 1998; Nozette et al., 1994). Global models of lunar topography have improved considerably in terms of spatial resolution and accuracy, when precise laser altimetry-based maps became available on Clementine (Zuber et al., 1994), Kaguya (SELENE) (Araki et al., 2009), and Lunar Reconnaissance Orbiter (Scholten et al., 2012; Smith et al., 2010). From Doppler tracking of the LP (Konopliv et al., 2001), 4-way Doppler tracking of the Kaguya (SELENE) spacecraft (Namiki et al., 2009), and spacecraft-to-spacecraft tracking observations of the dual Gravity Recovery and Interior Laboratory (GRAIL) spacecraft (Zuber et al., 2013), the global gravitational field of the Moon has been recovered with steadily increasing spatial and temporal resolution.

From LP Doppler data, Konopliv et al. (1998, 2001) had obtained basic information on the gravitational field of the Moon and improved values of the Moon's mass ($GM = 4902.801076 \pm 0.000081$ km^3 s^{-2}, with G the gravitational constant), mean MoI factor ($I/MR^2 = 0.3931 \pm 0.0002$), and tidal potential Love number ($k_2 = 0.026 \pm 0.003$). Based on a combined analysis of satellite tracking data from Lunar Orbiters I–V, the Apollo 15 and 16 subsatellites, Clementine, LP, and some SMART-1 spacecraft tracking data, Goossens and Matsumoto (2008) derived a lower value of $k_2 = 0.0213 \pm 0.0075$, in good agreement with determinations from lunar laser ranging. The tidal response of the Moon indicates that the deep lunar interior is still sufficiently hot to retain partially molten rock and a fluid core. However, prior to Kaguya (SELENE), the resolution of global gravity field solutions had been limited by the lack of tracking data within 60° of the center of the lunar farside. The first gravity maps revealing the unknown subsurface structures of the lunar farside were accomplished by using two satellites in coplanar orbits, with one satellite placed in a high orbit as a relay satellite between the main orbiter and the Earth and a VLBI subsatellite for differential tracking of the relay satellite (Namiki et al., 2009). The 4-way Doppler measurements of Kaguya (SELENE) allowed an estimation of the global medium- to-short-scale gravitational field of the Moon up to degree 70 without invoking unproven a priori constraints (Matsumoto et al., 2010).

The GRAIL primary mission has improved globally the precision of the gravity spectrum by up to three or four orders of magnitude since previous gravity field determinations were mostly limited to the lunar nearside (Konopliv et al., 2013; Lemoine et al., 2013). As a consequence, present knowledge of the lunar low-degree field has improved considerably and factors directly into more accurate solutions for the degree-2 and degree-3 tidal potential Love numbers k_2 and k_3. Whereas Konopliv et al. (2013) deduced $k_2 = 0.02405 \pm 0.00018$ and $k_3 = 0.0089 \pm 0.0021$, Lemoine et al. (2013) determined $k_2 = 0.024615 \pm 0.0000914$ and a preliminary value of $k_3 = 0.00734 \pm 0.0015$, illustrating the present level of inherent uncertainty of the different gravity estimation techniques applied.

Lunar laser ranging data show that the true spin axis of the Moon is displaced from the Cassini alignment (mean direction of the spin axis) by 0.26 arcsec, possibly caused by internal dissipation in the presence of partially molten rock and a fluid core (Dickey et al., 1994; Williams et al., 2001; Yoder, 1981). Magnetic field measurements were made by small satellites left in lunar orbit by the Apollo 15 and 16 missions. Although localized regions of magnetized rock were detected by the subsatellites, no global lunar magnetic field could be measured. A lunar magnetic dipole moment can be no larger than 10^{16} A m^2, that is, about seven orders of magnitude smaller than the Earth's dipole moment. The absence of a present-day global lunar magnetic field can be explained by the absence of an active lunar core dynamo. However, the paleomagnetic record of some lunar samples suggests the former existence of a significant magnetic field produced by dynamo action in a liquid metallic core (Hood and Jones, 1987). A convectively driven lunar dynamo might not be expected to last more than a few hundred million years until convection in the core has ceased. If the lunar dynamo lasted for a longer period of time, other mechanisms, such as precession (Dwyer et al., 2011) and compositional convection due to inner core freeze out, must be invoked to explain long-lived dynamo action. The recent detection of strong magnetization in a mare basalt sample indeed suggests the former existence of a relatively long-lived lunar core dynamo that may have operated between 4.2 and 3.7 billion years ago (Shea et al., 2012).

The Clementine topographic and gravity maps of Zuber et al. (1994) were the first reliable global characterizations of surface elevations and the gravitational field of the Moon. On the basis of those data, the lunar highlands appeared to be nearly isostatically compensated; however, basin structures showed a wide range of compensation states independent of their sizes or ages (Zuber et al., 1994). Beneath all resolvable basin structures, Zuber et al. (1994) observed crustal thinning and concluded that the structure and thermal history of the Moon are more complex than previously assumed. The GRAIL tracking observations have allowed the construction of the global gravitational field up to degree and order 660 (Konopliv et al., 2013; Lemoine et al., 2013), revealing a number of tectonic-, volcanic-, and impact-related geologic features not resolved in previous gravity maps. In particular, topographic features preserved in the highly fractured lunar crust apparently predominate the gravitational signature in the mid- to high-degree field up to degree and order 420 (Zuber et al., 2013). Using gravity gradiometry on the GRAIL observations, a globe-encircling population of long and elongated positive gravity anomalies has been detected, which are interpreted in terms of vertical sheetlike intrusions of solidified melt at depth indicative of lithospheric extension. The distribution, orientation, size, and stratigraphy of these linear subsurface features or dikes suggest that early in lunar history, the Moon experienced a state of thermally induced global extension by 0.6–4.9 km in radius (Andrews-Hanna et al., 2013). Therefore, the new high-resolution gravitational field data are also informative of improved thermal evolution models of the Moon.

Previous estimates of the crustal thickness varied from 30 to 35 km under basins associated with mass concentrations (mascon basins) to 90–110 km beneath the highlands, whereas irregular maria have intermediate thickness values of 50–60 km (Bills and Ferrari, 1977). From a reanalysis of gravity data acquired by the Apollo and Clementine spacecraft and assuming a thickness of 55 km at the Apollo 12/14 site, the thickness of the lunar farside crust has been estimated at 67 km, resulting in a mean thickness of 61 km if a uniform

crustal composition is assumed (Neumann et al., 1996). Analysis of global gravity and topography data implies that the thickness of the farside crust exceeds that of the nearside by about 15 km on average (Wieczorek et al., 2006). Nevertheless, the gravity-based results are model-dependent, and it is possible to construct models without large differences in crustal thickness between the nearside and the farside hemispheres. A difference in crust thickness between nearside and farside is thought to contribute substantially to the 1.7 km offset of the center of mass relative to the center of figure of the Moon in the direction of Earth (Hood and Jones, 1987; Kaula et al., 1974; Zuber et al., 1994). Wieczorek and Phillips (1998) had computed a variety of crustal thickness maps for the Moon, assuming both homogeneous and dual-layered crusts. The homogeneous crust model characterized by a constant crustal density provides a total crust thickness of 66 km, whereas the preferred dual-layered model of the lunar crust consisting of a 31 km thick upper crust and a 29 km thick lower crust yields a total thickness of 60 km. The gravitational field measurements of the LP spacecraft suggest a mean crustal thickness of about 70 km assuming Airy compensation of the lunar highlands (Konopliv et al., 1998) (see **Chapter 10.05** in this volume).

Early seismic studies favored crustal thickness of about 60 km beneath the Apollo network (Goins et al., 1981; Nakamura et al., 1982; Toksöz et al., 1974). An early analysis of body wave phases from artificial impacts of known impact time and location provided evidence for a dual-layered crust about 60 km thick beneath the Apollo 12 and 14 stations near Mare Cognitum (Toksöz et al., 1974), whereas a putative thickness of 75 ± 5 km was derived from seismic data near the Apollo 16 highland site (Goins et al., 1981). More recent studies, however, favor thinner crusts of only 45 km (Khan et al., 2000), 38 km (Khan and Mosegaard, 2002), or 30 km (Gagnepain-Beyneix et al., 2006). Chenet et al. (2006) attributed travel time variations observed from impact events to crustal thickness variations at impact sites and obtained thicknesses between 30 and 40 km at most sites that are also compatible with gravity field data. The mean thickness of the lunar crust is estimated as 49 ± 15 km if an Airy-type compensation mechanism (hydrostatic balance between crust and more dense mantle) applies (Wieczorek et al., 2006). A bulk density of the lunar highlands crust of 2550 kg m^{-3}, substantially lower than previously thought, has been inferred from the GRAIL high-resolution gravity data and suggests an average crustal porosity of 12% to depths of at least a few kilometers (Wieczorek et al., 2013). Furthermore, these authors find that the low bulk crustal density can be used to construct a global crustal thickness model in agreement with the Apollo seismic constraints and that the bulk refractory element composition of the Moon is not required to be enriched relative to that of Earth in case of a thin lunar crust with a mean thickness between 34 and 43 km.

Whereas early determinations of the MoI factor (Konopliv et al., 1998) are consistent with a core radius between 220 and 450 km, independent observations of the lunar magnetic moment induced in the geomagnetic tail of the Earth suggest that the lunar core radius is 340 ± 90 km (Hood et al., 1999). Using electromagnetic sounding methods, Shimizu et al. (2013) deduced estimates of the electrical conductivity of the lunar mantle and the size of the lunar core from Kaguya (SELENE) magnetic field data. The core radius is related to the ratio between the induced and inducing magnetic dipole moments and estimated to be 290 km with an upper bound of about 400 km (95% confidence limit) (Shimizu et al., 2013). Based on the knowledge of the Moon's mean density and MoI factor as derived from LP data and the tidal potential Love number k_2 and quality factor Q as obtained from the entire set of lunar laser ranging data, Khan et al. (2004) obtained information on the radial distributions of density and seismic velocities. The resultant structural model suggests the presence of a molten or partially molten iron core with radius and density of about 350 km and 7200 kg m^{-3}, respectively (Khan et al., 2004). A recent reprocessing of the Apollo seismic data that once allowed the existence of a lunar core with a radius of 170–360 km (Nakamura et al., 1974) supports not only the presence of a partially molten mantle layer above a liquid core with a radius of 330 km but also the possible existence of a small solid inner core (Garcia et al., 2011; Weber et al., 2011). The possible range of density models that simultaneously satisfy the mean density and MoI of the Moon is shown in **Figure 7**.

10.02.6.3 Composition

The lunar surface is divided into light-colored heavily cratered highlands and smooth dark lowland maria, which are most prominent on the nearside. The circular maria are frequently associated with mass concentrations (mascon basins) causing a positive anomaly in the lunar gravity field due to the larger density of the basalt layer compared to that of the surrounding anorthositic crust. Uplift of the crust–mantle interface provides another contribution to the positive gravity anomaly. Some mascon basins that do not appear to be associated with mare volcanism are known (Konopliv et al., 1998, 2001). Neumann et al. (1996) and Wieczorek and Phillips (1999) had suggested that the lunar mascons might partly result from superisostatic uplift of the crust–mantle interface. The highlands are

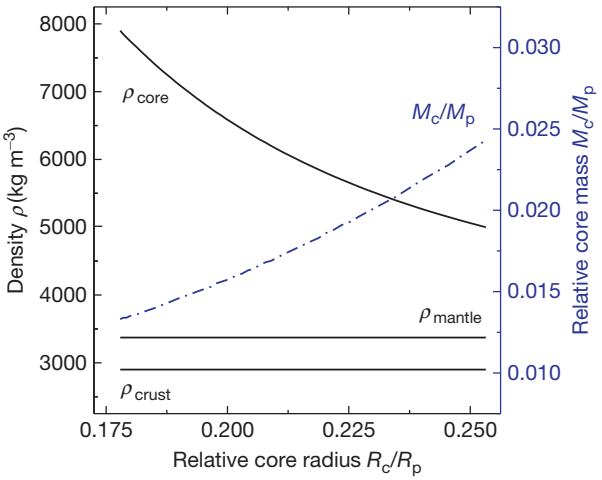


Figure 7 Three-layer model determinations of lunar mantle and core density and core mass fraction M_c/M_p versus relative core radius R_c/R_p consistent with the mean density and moment of inertia factor of the Moon (MoI = 0.3931). The thickness and density of the anorthositic lunar crust are fixed at 60 km and 2900 kg m^{-3}, respectively.

saturated with large craters owing to their greater age in comparison with the maria and dominate the lunar farside and most of the nearside. Highly anorthositic rocks are exposed in the lunar highlands, whereas the maria represent floods of basaltic lava that were erupted about ~0.5 Gy after major impact events. Little mare volcanism is associated with the South Pole–Aitken basin, and Oceanus Procellarum, the largest expanse of mare volcanism, is not necessarily associated with an impact event (Neumann et al., 1996).

The range of chlorine isotope compositions of Apollo basalts and glasses is attributed to volatile release of metal halides during volcanic eruptions, suggesting that the lunar interior is almost anhydrous with hydrogen concentrations reduced by a factor of $\sim 10^4$–10^5 as compared to Earth's (Sharp et al., 2010). This is consistent with a collisional origin of the Moon that predicates a substantial lack of water or hydrogen in the lunar interior (Canup and Asphaug, 2001). This view has been challenged, however, by the recent detection of mineral-bound water by applying highly sensitive analytic techniques to volcanic glasses (Hauri et al., 2011; Saal et al., 2008) and OH^+-bearing apatite minerals in lunar samples to infer the hydrogen content and isotopic composition in terms of D/H ratios (Boyce et al., 2010; Greenwood et al., 2011; McCubbin et al., 2010b). These estimates range from (i) a lunar mantle water content similar to that of the Earth's mantle; to (ii) 'wet' conditions, but at much lower water abundances, and D/H ratios compliant with some cometary water present in the lunar interior; to (iii) terrestrial-like original values subject to fractionation due to outgassing and thereby resulting in elevated D/H ratios (e.g., Jaumann et al., 2012, and references therein). In the absence of lithospheric plate recycling, however, it is difficult to explain the entrainment of cometary water into the lunar mantle. The issue of the volatile inventory of the lunar interior remains unresolved at this point but will benefit from steadily improving laboratory instrumentation and analytic techniques.

It appears that the majority of the lunar basalts erupted within the Procellarum KREEP Terrane (Jolliff et al., 2000a), a unique geologic province that contains elevated abundances of heat-producing elements. Wieczorek and Phillips (2000) had shown that the enhanced heat production of this province could have melted the underlying mantle. Wieczorek et al. (2001) had suggested that mare basalts might preferentially erupt within the impact basins since mare basalts are more dense than the upper anorthositic crust but less dense than the deep, more mafic, lower crust. If an impact event stripped away the upper crust, mare basalts could easily rise through the crust and erupt based solely on buoyancy considerations. The compositions of the mare basalts are consistent with volcanic source regions of several hundred kilometers in depth (Heiken et al., 1991). Mare basalt samples provide an assessment of the Mg number MgO/(MgO + FeO) of 0.75–0.8 and a bulk Al_2O_3 content of <1 wt.% for the upper mantle (Hood and Zuber, 2000). The ages of the basaltic maria, as determined from returned samples and careful crater chronological studies, range between about 4 and ~1.2 Gy (Hiesinger et al., 2003) and suggest that U-, Th-, and K-rich residua of the crystallized magma ocean providing the heat required for mantle melting were at least locally present (Jolliff et al., 2000a; Wieczorek and Phillips, 2000). The lateral and depth variations of crustal composition have been assessed from compositional mapping of central peaks (Wieczorek and Zuber, 2001) and ejecta blankets of large impact basins (Bussey and Spudis, 2000). Deeply excavated ejecta are found to be more mafic than the surface material (Pieters and Tompkins, 1999).

10.02.6.4 Mineralogy

The Moon is the only planetary body other than the Earth for which a seismic velocity structure has been derived from analyses of the Apollo seismic data set (see **Chapter 10.03** in this volume). A seismic network of four stations on the lunar nearside, installed at the landing sites of Apollos 12, 14, 15, and 16 from 1969 to 1972, continued to operate until it was turned off in September 1977. With few exceptions, almost all seismic events within the Moon occurred on the nearside (Nakamura, 2005). The seismic activity of the Moon is provided by deep moonquakes located at 850–1000 km depth, shallow moonquakes, or high-frequency teleseismic events situated at 50–220 km depth and meteoroid impacts onto the lunar surface. The deep moonquakes, apparently confined to the lunar nearside, are most numerous and triggered by tidal deformation of the Moon. More than 12 000 events associated with 81 identified sources ('nests') were recorded during the lifetime of the Apollo seismic network. During the limited operational period of the lunar seismic network with stations on the near-Earth hemisphere, only one meteoroid event was detected with seismic rays crossing a central low-velocity zone from which the size of a lunar core was estimated for the first time. Early seismic analyses were based on a limited set of arrival time readings (Goins et al., 1981), whereas the model of Nakamura (1983) employs the complete data set of 5-year simultaneous operation of four Apollo seismometers including more deep moonquake sources. More detailed lunar velocity and density structure models have been obtained from a new inversion of different subsets of the Apollo seismic data set by applying various inversion techniques to revised arrival time readings and seismological array techniques (Chenet et al., 2006; Garcia et al., 2011; Khan and Mosegaard, 2002; Khan et al., 2000, 2004; Lognonné et al., 2003; Weber et al., 2011).

The Apollo seismic data record indicates that the chemical composition of the silicate mantle is consistent with that of an olivine–pyroxene mixture (Hood and Jones, 1987; Toksöz et al., 1974). The seismic velocities are almost constant in the upper mantle and more heterogeneously distributed in the middle mantle below. The most prominent increase in seismic velocity at a depth of ~500 km below the surface of the Moon represents the transition from the upper mantle to the middle mantle (**Figure 8**). Nevertheless, recent analyses do not require a discontinuity at this depth (Khan et al., 2007). If this discontinuity exists and is global in extent, it may represent the initial depth of melting and differentiation during the magma-ocean phase of the Moon. If it is only present on the nearside hemisphere, it might instead reflect the maximum depth of melting of the mare basalt source region underneath the Procellarum KREEP Terrane (Wieczorek and Phillips, 2000).

This discontinuity has been interpreted in terms of a mineralogical phase transition from the spinel to the garnet stability field. Alternatively, a change in composition to more aluminous and MgO-rich mafic silicates has been invoked,

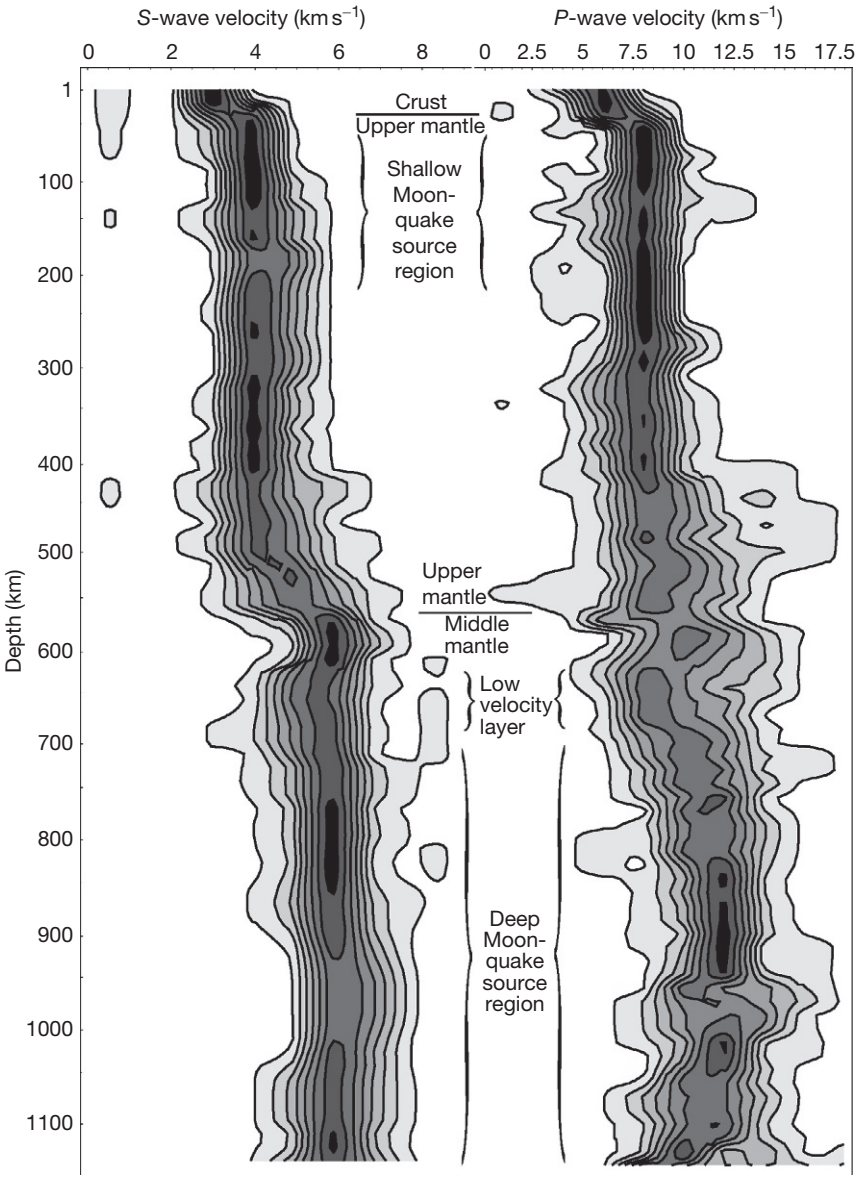


Figure 8 Marginal posterior probability distributions illustrating the range of possible (left) *S*-wave and (right) *P*-wave velocity structures of the Moon based on a total number of 50 000 models. The contour lines define nine equally sized probability density intervals for the distributions. Note that maximum probability does not necessarily correspond to maximum likelihood of a calculated velocity. Adapted from Khan A and Mosegaard K (2002) An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data. *Journal of Geophysical Research* 107(E6), http://dx.doi.org/10.1029/2001JE001658, with permission American Geophysical Union; Copyright 2002.

thereby increasing the Mg number below this depth (Khan et al., 2006; Nakamura, 1983). The reanalysis of the Apollo lunar seismic data indicates a homogeneous, constant-velocity upper mantle extending down to 560 ± 15 km depth, whereas the radial velocity distribution suggests more inhomogeneous middle and lower mantle layering (Khan and Mosegaard, 2002; Khan et al., 2000). The compositional change across the discontinuity together with the homogeneity of the upper mantle and the inhomogeneity of the middle mantle has been interpreted in terms of the initial depth of melting and differentiation during the magma-ocean phase of the Moon (Hood and Zuber, 2000). Based on mare basalt petrology and thermal evolution considerations, Elkins-Tanton et al. (2003a) identified the 500 km discontinuity with the maximum depth of melting beneath the Procellarum KREEP Terrane on the lunar nearside. It is also possible that the 500 km discontinuity represents stratified olivine- and orthopyroxene-rich cumulates that were subsequently emplaced at the bottom of the lunar magma ocean. The mineralogical layering is supported by seismic inversions that account for thermodynamic mineral phase equilibriums (Kuskov, 1995, 1997; Kuskov and Fabrichnaya, 1995; Kuskov and Kronrod, 1998).

In the lower mantle below a depth of about 1150 km, seismic waves are strongly attenuated. This has been interpreted in terms of a partial melt zone that extends down to the core–mantle boundary (Nakamura, 2005; Weber et al.,

2011). Because of the depth distribution of the deep moonquakes, the *P*- and *S*-waves do not sample depths below 1200 km, and thus, the core region remains poorly understood. Only one among the ~1700 meteoroid impacts occurred sufficiently distant to facilitate investigations of the deepest lunar mantle. This observation constrained the lunar core radius to a value between 170 and 360 km (Nakamura et al., 1974). Weber et al. (2011) used seismological array techniques to identify reflected phases from the deep interior of the Moon not resolved previously. These results are consistent with the existence of a partial melt layer in the deep mantle, a liquid iron-rich core, and a solid inner core, commensurate with previous studies using different data sets. The thicknesses of these layers, but not the overall structure, are subject to inherent uncertainties connected to the poorly known mantle structure, since the travel times of the new seismic phases are sensitive to the average propagation velocities within the mantle. The radius of the outer core has also been constrained by the detection of ScS phases in a set of carefully chosen recordings from sources with favorable orientation (Garcia et al., 2011).

10.02.6.5 Future Exploration

Considering the ambiguity inherent in the Apollo lunar seismic data, first reliable recordings of seismic phases traveling through the central region of the Moon must be expected from future landed missions. Models of the internal structure of the Moon would greatly benefit from the joint analysis of the lunar gravity field, topography, rotational state, tidal distortion, and orbital parameters combined with inversions of the entire record of lunar laser ranging, magnetic induction, and seismic data (Wieczorek et al., 2006; Williams, 2007). There are many open questions regarding the present constitution, origin, and evolution of the Moon that would be best addressed in the frame of a geophysical network mission (Mocquet et al., 2011). A higher temporal resolution and precision compared to ground- and orbit-based techniques alone can be achieved by emplacement of a geophysical network of one or several soft-landed surface stations on the Moon. The Apollo legacy suggests that geophysical network data could help improve the present knowledge of the Moon's interior structure, the lunar environment, and the dynamics of the Earth–Moon system.

10.02.7 Mercury

The planet Mercury is unique among the terrestrial planets in many respects. It represents an end-member of the terrestrial planets with respect to its density and distance from the Sun and thereby provides important constraints on planetary formation and evolution in the innermost part of the solar nebula (Balogh and Giampieri, 2002; Solomon, 2003; Strom, 1987).

10.02.7.1 General

Mercury is 4880 km in diameter, roughly 1/3 the size of Earth, and occupies only about 6% of the volume of Earth. Albeit substantially smaller in size, Mercury's surface gravity of 3.7 m s^{-2} is like that of the larger planet Mars. In 1974 and 1975, the Mariner 10 spacecraft provided the first close-up look of the planet during three encounters. From these flybys, better values of the planet's mass, mean radius, and average density were obtained. Less than half of the surface was covered by images at an average resolution of 1 km and < 1% at 100–500 m resolution, resulting in a limited characterization of surface morphology and geologic evolution. Among the most important findings of the Mariner 10 spacecraft was the then unexpected detection of a magnetic field of internal origin (Ness, 1979).

Our knowledge of Mercury from the Mariner flybys of the mid-1970s has now been superseded by newer observations of the planet, especially by the MESSENGER spacecraft that went into orbit around Mercury on 18 March 2011. Prior to orbit insertion, MESSENGER flew by Mercury three times and each of these flybys contributed substantially to our knowledge of the planet. Even before MESSENGER, our understanding of Mercury took a large step forward when radar observations of its surface from Earth succeeded in measuring Mercury's obliquity and the relatively large amplitude of the longitudinal librations of its rotation axis. These measurements, when combined with the Mariner 10 determination of Mercury's gravitational coefficient $C_{2,2}$, required that Mercury's core and mantle are decoupled by at least a partially molten core. The existence of a partially molten core lends strong support to the hypothesis that Mercury's magnetic field is produced by an active core dynamo.

The surface of Mercury is covered by densely cratered regions, intercrater plains, and smooth plains. The heavily cratered areas are some of the oldest surfaces in the solar system. Intercrater and smooth plains were likely emplaced volcanically though some could have formed of impact melts or other basin ejecta that behaved fluidlike upon deposition (Wilhelms, 1976). Lobate scarps and widespread volcanic plains suggest an early evolution in which volcanically induced expansion preceded a phase of planetary contraction (Solomon, 2003). Observations of Mercury's surface by the MESSENGER spacecraft have extended the Mariner 10 coverage to essentially the entire planet and provide a global view of Mercury's impact, tectonic, and volcanic history. Crater densities suggest that the smooth plains within and around the Caloris Basin, the largest impact basin on Mercury, are younger than intercrater plains elsewhere (Fassett et al., 2011; Strom et al., 2011). Even the most densely cratered regions on Mercury are deficient in craters between 20 and 100 km in diameter compared with the Moon, possibly as a consequence of early volcanic resurfacing (Fassett et al., 2011). For craters larger than 100 km in diameter, crater density on Mercury in a given size range is similar to that of the Moon.

Lobate scarps are the dominant tectonic landform on Mercury, and according to MESSENGER observations, they record a total contractional strain of at least one-third greater than inferred from Mariner 10 images (Solomon et al., 2008). The scarps must have formed at the end of or subsequent to the early phase of severe bombardment on Mercury because they sometimes pass through and deform old craters. The scarps may be a consequence of the cooling and contraction of the core, and if so, they are unique surface features that distinguish Mercury with its large core from the Moon with only a very small core.

MESSENGER observations have revealed that volcanism has played an important role in shaping Mercury's surface (Head et al., 2011). A large contiguous expanse of smooth plains covers much of Mercury's high northern latitudes. More than 6% of Mercury's surface is blanketed by these smooth volcanic plains. They embay other landforms, are distinct in color, and show several flow features. They partially or completely bury impact craters, whose sizes indicate plains thicknesses of more than 1 km and multiple phases of emplacement. These characteristics suggest emplacement in a flood-basalt style, consistent with x-ray spectrometric data indicating surface compositions intermediate between those of basalts and komatiites. Crater densities on the plains show that they formed after the Caloris impact basin, confirming that volcanism was a globally extensive process in Mercury's post-heavy bombardment era (Head et al., 2011).

High-resolution MESSENGER images of Mercury's surface reveal that many bright deposits within impact craters show fresh-appearing, irregular, shallow, rimless depressions or hollows (Blewett et al., 2011). The hollows are tens of meters to a few kilometers across, and many have high-reflectance interiors and halos. The host rocks are associated with crater central peaks, peak rings, floors, and walls. The most likely formation mechanisms for the hollows involve recent loss of volatiles through some combination of sublimation, space weathering, outgassing, or pyroclastic volcanism (Blewett et al., 2011). These features support the inference that Mercury's interior contains higher abundances of volatile materials than predicted by most scenarios for the formation of the planet (Blewett et al., 2011).

The neutron spectrometer (Lawrence et al., 2013) and the laser altimeter (Neumann et al., 2013) instruments on the MESSENGER spacecraft have found evidence of water ice and other deposits in permanently shadowed craters at Mercury's north pole. The first indication that there might be ice at Mercury's poles came from Earth-based measurements of radar-bright regions near Mercury's north and south poles (Slade et al., 1992). Subsequent measurements showed that these unusual radar characteristics are confined to permanently shadowed regions within high-latitude impact craters (Harmon et al., 2011). The leading explanation for the high radar reflectance is the presence of large amounts of water ice that can be thermally stable in regions of permanent shadow over geologically long periods of time (Harmon et al., 2011). The MESSENGER neutron spectrometer measured enhanced concentrations of hydrogen, both at the surface in permanently shadowed craters and sometimes buried beneath other deposits, consistent with pure water ice. The MESSENGER laser altimeter observed surface reflectance anomalies over permanently shadowed poleward-facing slopes near Mercury's north pole that are spatially collocated with areas of high radar backscatter thought to be due to near-surface water ice. The total mass of water at Mercury's poles is inferred to be 2×10^{13}–10^{15} kg and is consistent with delivery by comets or volatile-rich asteroids (Lawrence et al., 2013). Both Mercury and the Moon (Mitrofanov et al., 2010) have preserved water ice at their poles.

The major element composition of Mercury's surface has been measured by the x-ray spectrometer on MESSENGER (Nittler et al., 2011). The x-ray fluorescence spectra show that Mercury's surface has a different composition from the surfaces of other terrestrial planets. Relatively high Mg/Si and low Al/Si and Ca/Si ratios rule out a lunar-like feldspar-rich crust. Mercury's ratios are intermediate between typical basaltic compositions and more ultramafic compositions comparable to terrestrial komatiites (Nittler et al., 2011). The sulfur abundance is at least ten times higher than that of the silicate portion of the Earth or the Moon, and this observation, together with a low surface Fe abundance, supports the view that Mercury formed from highly reduced precursor materials, similar perhaps to enstatite chondrite meteorites or anhydrous cometary dust particles (Nittler et al., 2011), and the planet's core might contain Si as well as sulfur (Malavergne et al., 2010).

The MESSENGER gamma ray spectrometer measured the average surface abundances of the radioactive elements potassium (K, 1150 ± 220 ppm), thorium (Th, 220 ± 60 parts per billion), and uranium (U, 90 ± 20 parts per billion) in Mercury's northern hemisphere (Peplowski et al., 2011). The abundance of the moderately volatile element K, relative to Th and U, is inconsistent with physical models for the formation of Mercury requiring extreme heating of the planet or its precursor materials. Processes involving high temperatures such as vaporization of an early mantle and crust or a giant impact have been proposed to explain the large metal-to-silicate ratio of Mercury. Abundances of K, Th, and U indicate that internal heat production has declined substantially since Mercury's formation, consistent with widespread volcanism shortly after the end of late heavy bombardment 3.8 billion years ago and limited, isolated volcanic activity since (Peplowski et al., 2011).

Mercury's orbit about the Sun has a semimajor axis of 0.387 AU, an eccentricity of 0.206, and an inclination of 7° relative to the ecliptic. Therefore, the orbital distance from the Sun varies between 0.308 AU at perihelion and 0.466 AU at aphelion in the course of one revolution, causing significant tidal distortion of the planet (Burns, 1976; Van Hoolst and Jacobs, 2003). Furthermore, the near-surface layer of Mercury is exposed to a severe thermal environment in terms of elevated surface temperatures and high subsurface temperature gradients due to the planet's proximity to the Sun with an insolation of up to 15 kW m^{-2}. Related diurnal surface temperature variations between 90 and 740 K are greater than on any other planet or satellite in the solar system (Strom, 1987, 1997). Nevertheless, there are permanently shadowed regions at the bottoms of some craters at Mercury's poles where temperatures remain sufficiently low throughout a Mercurian year to insure the long-term stability of water ice (Paige et al., 2013).

Radar observations from the Earth have shown that the rotation period (58.6 days) is locked into a 3:2 resonance with the orbital period (87.7 days). As a consequence of the 3:2 spin–orbit coupling caused by tidal interactions with the Sun, one solar day on Mercury lasts 176 Earth days and corresponds to two revolutions or three rotations of the planet. While the distribution of solar irradiation is symmetrical between both hemispheres, the significant orbital eccentricity causes longitudinal variations superimposed on the latitudinal variation of the solar irradiation (Van Hemelrijck and Vercheval, 1981).

Owing to the lack of a substantial atmosphere, the surface of Mercury has been heavily cratered and fragmented by

impacts and small-particle bombardment resulting in a planet-wide, fine-grained, and thermally insulating regolith layer (Langevin, 1997). The thermal and electrical properties of the Mercurian regolith derived from Mariner 10 measurements are similar to those of the Moon (Chase et al., 1976). Even large-amplitude surface temperature perturbations are thus expected to rapidly fade away with depth and will be negligible below several thermal skin depths (Snyder Hale and Hapke, 2002; Vasavada et al., 1999).

10.02.7.2 Interior Structure

The large average density of $5430 \pm 10\ \mathrm{kg\ m^{-3}}$ (Anderson et al., 1987) is comparable to that of the Earth and Venus but much larger than that of the Moon and Mars. The corresponding zero-pressure density of about $5300\ \mathrm{kg\ m^{-3}}$ is even much higher than the uncompressed densities of Earth, Venus, and Mars, which are about 4100, 4000, and $3800\ \mathrm{kg\ m^{-3}}$, respectively. This suggests that Mercury contains a larger proportion of heavier elements such as iron than any other terrestrial planet, which is concentrated in a substantial core. The freezing of an inner pure iron core would be accompanied by the enrichment of light elements such as sulfur in a liquid outer core (**Figure 9**) and has implications for the concentration of mass toward the center. The mass concentration of iron should be about twice that in the Earth (Wasson, 1988).

Radio tracking of the MESSENGER spacecraft has allowed the determination of the low-degree and low-order spherical harmonic coefficients of Mercury's gravity field (Smith et al., 2012). Values of GM, $C_{2,0}$, and $C_{2,2}$ are, respectively, $22031.84\ \mathrm{km^3\ s^{-2}}$, -2.252×10^{-5}, and 1.253×10^{-5}. It is possible to infer Mercury's MoI from the gravitational and rotational parameters because of the planet's special spin state. Mercury is in a Cassini state wherein the axis of rotation, the orbit normal, and the normal to the Laplace plane remain coplanar as the spin vector and the orbit normal precess together about the latter with an ~300 000-year period (Margot et al., 2007).

As already mentioned, Earth-based radar observations also show that the amplitude of the 88-day forced libration in longitude is so large that the mantle and crust are librating independently of the core (Margot et al., 2007). This dynamic state together with the values of $C_{2,0}$ and $C_{2,2}$, the Earth-based radar measurements of the amplitude of Mercury's forced libration (35.8 ± 2 arcsec) and obliquity (2.06 ± 0.1 arcmin), and data on the precession rate and pole position provide the basis for calculating not only C/MR^2 but also C_m/C, where C_m is the axial MoI of Mercury's silicate shell. A model for Mercury's radial density distribution based on these results includes a solid silicate crust and a rocky mantle overlying an iron-rich liquid outer core (**Figure 10**). Mercury could have a solid inner core. The outer radius of the core is inferred to be 2030 ± 37 km (Margot et al., 2007; Smith et al., 2012). Mercury's core, with a radius about 83% of the planet's radius, contains most of the planet's iron.

Hauck et al. (2013) constructed models of the radial density structure of Mercury based on the radius and density of the planet and the moments of inertia of the planet and its silicate shell. These models are based on updated values of obliquity 2.04 ± 0.08 arcmin, the amplitude of the longitudinal libration 38.5 ± 1.6 arcsec, MoI $C/MR^2 = 0.346 \pm 0.014$, and $C_m/C = 0.431 \pm 0.025$ (Margot et al., 2012). Hauck et al. (2013) estimated that the top of the liquid core is at a radius of 2020 ± 30 km consistent with the location inferred by Smith et al. (2012). Hauck et al. (2013) found that the mean density above this boundary is $3380 \pm 200\ \mathrm{kg\ m^{-3}}$, and the density below the boundary is $6980 \pm 280\ \mathrm{kg\ m^{-3}}$. They infer the thickness of the outer solid shell to be 420 ± 30 km. They

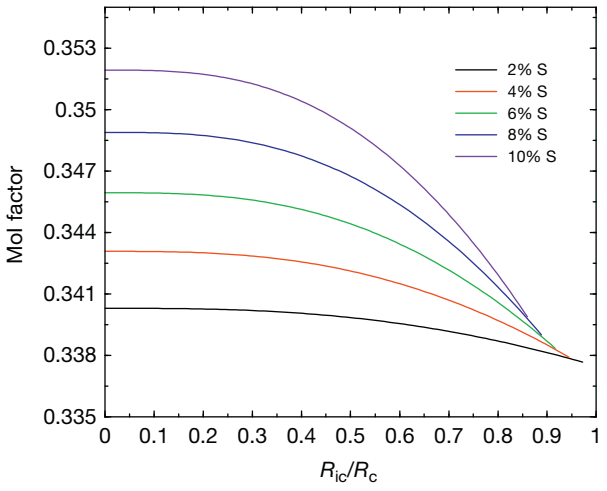


Figure 9 Mean moment of inertia (MoI) factor of a three-layer model of Mercury as a function of inner core radius R_{ic} and core chemistry. R_{ic} is given relative to the radius of the core R_c.

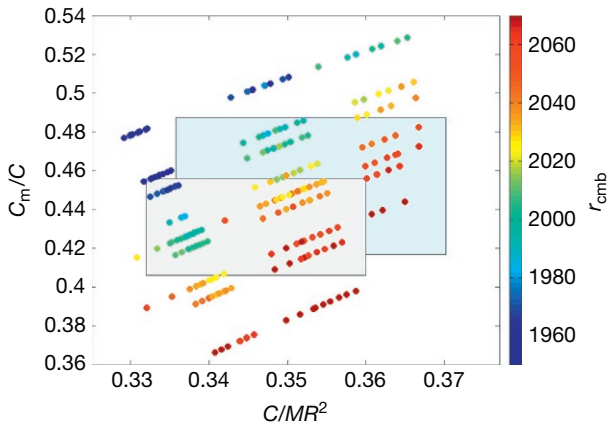


Figure 10 Model determinations of Mercury's core radius R_c (color-coded) as a function of the axial MoI factor, C/MR^2, and the ratio C_m/C between the axial MoI of the planet's solid portion and that of the entire planet. The latter depends on the coupling between core and mantle and provides clues to the physical state of the outermost part of Mercury's core. C_m/C would be around 0.5 for a liquid core or liquid outer core shell and about 1 in case of a solid core. Shaded boxes indicate the observational uncertainty range ($1-\sigma$) according to Smith et al. (2012) (turquoise) and Margot et al. (2012) (light gray). Reproduced from Steinke T (2012) *Modeling Mercury's Interior Structure and Tidal Deformation*. Bachelor's Thesis (in German), Karlsruhe Institute of Technology, Karlsruhe.

claim that these internal structure parameters are robust across a broad range of compositional models for the core and planet as a whole, a claim supported by their Monte Carlo internal structure modeling techniques.

There are several large gravity anomalies in Mercury's northern hemisphere with amplitude about 100 mGal. One of these anomalies coincides with the Caloris impact basin (Smith et al., 2012). The gravity field and the topography from MESSENGER's laser altimeter have been combined to produce a model of the crustal thickness of Mercury's northern hemisphere (Smith et al., 2012). The model assumes a crust–mantle density contrast of 200 kg m^{-3} and a mean crustal thickness of 50 km. The crust is generally thicker (50–80 km) near the equator and thins toward the north polar region (20–40 km); the regionally thinnest crust is located beneath the northern lowlands. The Caloris Basin overlies an area of locally thin crust, consistent with it being a mascon basin similar to those on the Moon.

Earth-based radar ranging data have suggested that the equatorial shape of Mercury has a significant ellipticity $(a–b)/a = (540 \pm 54) \times 10^{-6}$ and a center of figure CoF offset from the center of mass CoM of amplitude 640 ± 78 m in the approximate direction of the then unseen hemisphere (Anderson et al., 1996). MESSENGER laser altimeter measurements generally confirm these results, though there are some quantitative differences. The MESSENGER amplitude of the equatorial component of the CoF–CoM offset is 683 ± 146 m (Smith et al., 2010; Zuber et al., 2012) in approximately the same direction as the earlier radar ranging result. The equatorial shape of Mercury is not that of an equilibrium figure. The CoF–CoM offsets of all the terrestrial planets are plausibly attributed to hemispheric asymmetries in crustal thickness depending on the contrast $(\rho_m–\rho_s)/\rho_m$ between the densities of the mantle and the crust ρ_m and ρ_s, respectively. The magnitude of the CoF–CoM offset implies an excess crustal thickness of <12 km, which is comparable to that obtained for the Moon. Smith et al. (2010) inferred an even smaller crustal thickness difference. From a comparison between the equatorial shape and the gravitational equatorial ellipticity $C_{2,2}$ as inferred from the Mariner 10 flybys, Anderson et al. (1996) concluded that the Mercurian crust could be 200 ± 100 km thick if Mercury's equatorial ellipticity were entirely compensated by Airy isostasy (hydrostatic balance between crust and more dense mantle). Using the same approach, Smith et al. (2010) got a crustal thickness of 156 ± 23 km. They noted, however, that this is an overestimate since membrane stresses can contribute to the support of the equatorial ellipticity.

10.02.7.3 Composition

As discussed already in the preceding text, the major element surface composition of Mercury has been determined by the x-ray spectrometer on MESSENGER. The abundance of Fe in volcanic rocks at Mercury's surface is low. An upper bound on the average surface abundance of Fe of ~4 wt.% (Nittler et al., 2011) suggests that Mercury's silicate mantle is also low in iron. Most of Mercury's iron has segregated into its core. The low Fe abundance indicates that the surface *S* cannot be present primarily in the form of iron sulfides. More likely, the *S* occurs in Mg- and/or Ca-rich sulfides, which are stable under reducing conditions and present in highly reduced enstatite chondrite meteorites (Nittler et al., 2011). Relatively low surface abundances of Ti and Al (Nittler et al., 2011) suggest that Mercury's mantle contains limited amounts of high-density minerals as ilmenite and garnet.

It has been noted earlier that Mercury's relatively low Al/Si and Ca/Si ratios rule out the presence of a plagioclase-rich crust similar to that of the lunar highlands (Nittler et al., 2011). The Moon's ancient crust is believed to have formed by flotation of crystallized plagioclase-rich rocks that solidified from a global magma ocean. That such a crust never formed on Mercury is consistent with the low concentration of FeO in Mercury's mantle (Nittler et al., 2011). This does not rule out the possibility that Mercury had an early magma ocean. It only requires that upon solidification, the magma ocean, if it existed, did not form a plagioclase-rich, low-density crust.

If Mercury formed in a highly reducing environment as suggested by the low Fe and high S contents of its surface (McCubbin et al., 2012b; Nittler et al., 2011; Zolotov et al., 2013), then the planet's core might contain Si as well as sulfur (Malavergne et al., 2010). As already noted, Fe–S–Si alloys show liquid immiscibility at pressures <15 GPa and temperatures above the liquidus temperature (Morard and Katsura, 2010; Sanloup and Fei, 2004). Accordingly, FeS could have segregated and concentrated at the top of Mercury's core, perhaps forming a solid layer at the bottom of the silicate mantle (Hauck et al., 2013; Malavergne et al., 2010; Smith et al., 2012). In this case, the core–mantle boundary would not be a liquid–solid interface if the mantle is defined to be the silicate part of the planet.

However, the putative FeS layer is not a required feature of the acceptable density models of the planet. Using Bayesian inversion methods, Rivoldini and Van Hoolst (2013) calculated a broad range of compositionally distinct interior structure models of Mercury, assuming that sulfur is the only light element in the core. According to these models, Mercury has a core radius of 2004 ± 39 km, an average core density of 7233 ± 267 kg m^3, and a core sulfur content of 4.5 ± 1.8 wt. %. It is also concluded by these authors that the geodesy data impose strong constraints on the radius of the core and its composition in terms of average density or core sulfur content, whereas the mantle density is only weakly constrained by the geodesy data. Furthermore, Rivoldini and Van Hoolst (2013) pointed out that these models are not sufficiently diagnostic to distinguish between entirely liquid cores and the existence of a solid inner core.

10.02.7.4 Magnetic Field

Mercury is the only terrestrial planet other than Earth with an internally generated dipole-dominant magnetic field. The measured dipole moment is 190 nT R_M^3, where R_M is Mercury's mean radius (Anderson et al., 2011, 2012). The global planetary field is a southward-directed, spin-aligned, offset dipole centered on the spin axis. The offset is about 479 ± 6 km north of the geographic equator. The magnetic axis is tilted by $<0.8°$ from the rotation axis (Anderson et al., 2011, 2012). There is also evidence of crustal remanent magnetization on Mercury (Purucker et al., 2012). The northward offset of the dipole is equivalent to a strong quadrupole component described by the

Gauss coefficient $g_{2,0} = -74.6 \pm 4.0$ nT (Anderson et al., 2012). Mercury's magnetic field is unusual in the sense that multipolar components higher than degree 2 are small, while the dipole is dominant and the quadrupole is significant. Mercury's magnetic field is also weak compared with the Earth's field; Mercury's dipole moment is only about 10^{-3} that of the Earth's dipole moment.

There have been many attempts to explain the observed characteristics of Mercury's magnetic field. Thin-shell dynamos (Stanley et al., 2005; Takahashi and Matsushima, 2006); thick-shell dynamos (Heimpel et al., 2005); deep dynamos operating below a stably stratified, electrically conducting layer (Christensen, 2006; Christensen and Wicht, 2006; Manglik et al., 2010; Wicht et al., 2007); dynamos with induction feedback from magnetopause currents (Glassmeier et al., 2007a,b; Gómex-Pérez and Solomon, 2010; Gómex-Pérez and Wicht, 2010; Heyner et al., 2011a,b); and dynamos driven by precipitation of solid iron (Vilim et al., 2010) all can produce a weak field. However, thin-shell dynamos show a dipole offset much less than observed at Mercury or no quadrupole dominance over higher multipoles (Stanley et al., 2005). The feedback dynamo yields a planetary field that is axisymmetric but dominated by odd harmonics (Heyner et al., 2011b). Iron snow layers produce a field with a large octupole (degree 3) contribution and comparatively weak quadrupole structure (Vilim et al., 2010). Other dynamo models incorporate a sub-adiabatic heat flow at the core–mantle boundary to produce a stably stratified electrically conducting layer that filters out spherical harmonic terms of high degree and order that are the most variable in time. In these models, the field at the planetary surface is dominated by the axial dipole and quadrupole terms that are the most slowly time-varying components (Christensen, 2006; Christensen and Wicht, 2006; Stevenson, 1982; Wicht et al., 2007). In addition, zonal flows set up in the stably stratified layer preferentially reduce non-axisymmetric, low-degree terms (Christensen and Wicht, 2006), yielding a surface field that is weak, dipole-dominant with a substantial quadrupole component, and axisymmetric. The field structure produced by these models is highly dependent on the thickness of the stable layer and its dynamics (Manglik et al., 2010; Stanley and Mohammadi, 2008). It is not yet clear that we understand the processes behind the generation of Mercury's unusual magnetic field.

The presence of an internally generated magnetic field is consistent with an iron core that is at least partially liquid with an electrically conducting outer core of unknown thickness surrounding a solid inner core. Thermal evolution models indicate that the core would have solidified early in the history of Mercury unless a light alloying element such as sulfur was present. Even a small amount of sulfur is sufficient to depress the freezing point of a core alloy (Stevenson et al., 1983). The high surface abundance of sulfur suggests that despite its volatility and the closeness of Mercury to the Sun, sulfur is a likely constituent of the materials that formed Mercury and it therefore is present in Mercury's core. The enrichment of sulfur in the outer core upon freezing of an inner iron core would increase the depression of the freezing point and maintain a liquid outer core shell in spite of planet cooling (Figure 11). Thermal history models taking into account parameterized convective heat transport through the mantle indicate that bulk core sulfur concentrations of 1–5 wt.% are needed to retain a liquid outer core shell at the present time (Schubert et al., 1988; Spohn, 1991; Stevenson et al., 1983). Based on parameterized models of the coupled thermal, magmatic, and tectonic evolution of Mercury, Hauck et al. (2004) concluded that a dry-olivine mantle rheology, thorium-dominated radiogenic heating suggesting late silicate-mantle vaporization (Cameron et al., 1988), and a bulk core sulfur content of at least 6.5 wt.% are needed to explain both the planet's radial contraction of 1–2 km and the presence of a dynamo-driven, intrinsic magnetic field.

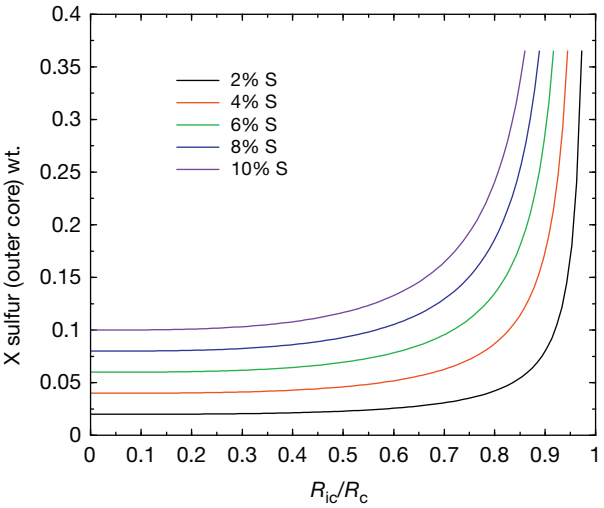


Figure 11 Sulfur content of the outer core of a three-layer model of Mercury as a function of inner core radius R_{ic} and core chemistry. R_{ic} is given relative to the radius of the core R_c.

Models of mantle convection including pressure- and temperature-dependent rheology demonstrate that the cooling history of a terrestrial planet is governed by the growth of its lithosphere, while the deep interior remains relatively hot. These models compare well to the parameterized convection calculations but produce thicker outer core shells at identical sulfur concentrations. Depending on the stiffness of the mantle rheology, a liquid outer core layer is then sustained even for sulfur concentrations as small as 0.2 wt.% consistent with cosmochemical arguments in favor of a volatile-poor planet (Spohn et al., 2001b). The 2 km radial contraction of Mercury in the absence of large-scale magmatism about 4 Gy may be linked to core shrinkage due to solid inner core growth and mantle cooling governed by lithospheric thickening and sluggish mantle convection (Schubert et al., 1988). If S and Si are present in Mercury's core, an FeS-rich layer could form at the top of the core and a part of it might be solid (Hauck et al., 2013; Smith et al., 2012). Such a layer could filter the dynamo-generated magnetic field as discussed earlier.

10.02.7.5 Future Exploration

Since Mercury is tidally flexed in its highly eccentric orbit about the Sun, the tidal Love numbers h_2 and k_2 can be derived from time-variable gravitational field measurements on a spacecraft orbiting the planet (Van Hoolst and Jacobs, 2003). The radial

displacement and tidal potential Love numbers h_2 and k_2, respectively, may provide useful constraints on the radial density and rigidity distribution and the extent of core differentiation for significant inner core sizes, that is, $R_i/R_c > 0.5$ (Steinke, 2012; **Figure 12**). Furthermore, the higher-order components of the gravitational field can be used to estimate crust thickness variations at short wavelengths and core–mantle boundary undulations at long wavelengths; the latter should be easily detectable due to the large size of Mercury's core relative to the planet's size (Spohn et al., 2001b).

No seismic data for Mercury exist. However, the periodic deformation of Mercury by solar tides may have important consequences for the planet's seismic environment. Owing to its highly elliptic orbit about the sun and its bound rotation, Mercury is exposed to strong tidal forces (Rivoldini et al., 2009; Van Hoolst and Jacobs, 2003). As a result, the seismic environment of Mercury may feature large numbers of lunar-like quakes.

10.02.8 Mars

The size of Mars is about half the size of the Earth and its mass is about 1/10 the mass of the Earth. The uncompressed density of roughly 3800 kg m^{-3} is significantly lower than the uncompressed densities of the Earth and Venus. The surface environmental conditions are the most Earthlike among the terrestrial planets with surface temperatures varying between 140 K at night in winter and 300 K at midday in summer. Mars is thought to be a one-plate planet, as are the Moon and Mercury, lacking plate tectonics at least at the present time.

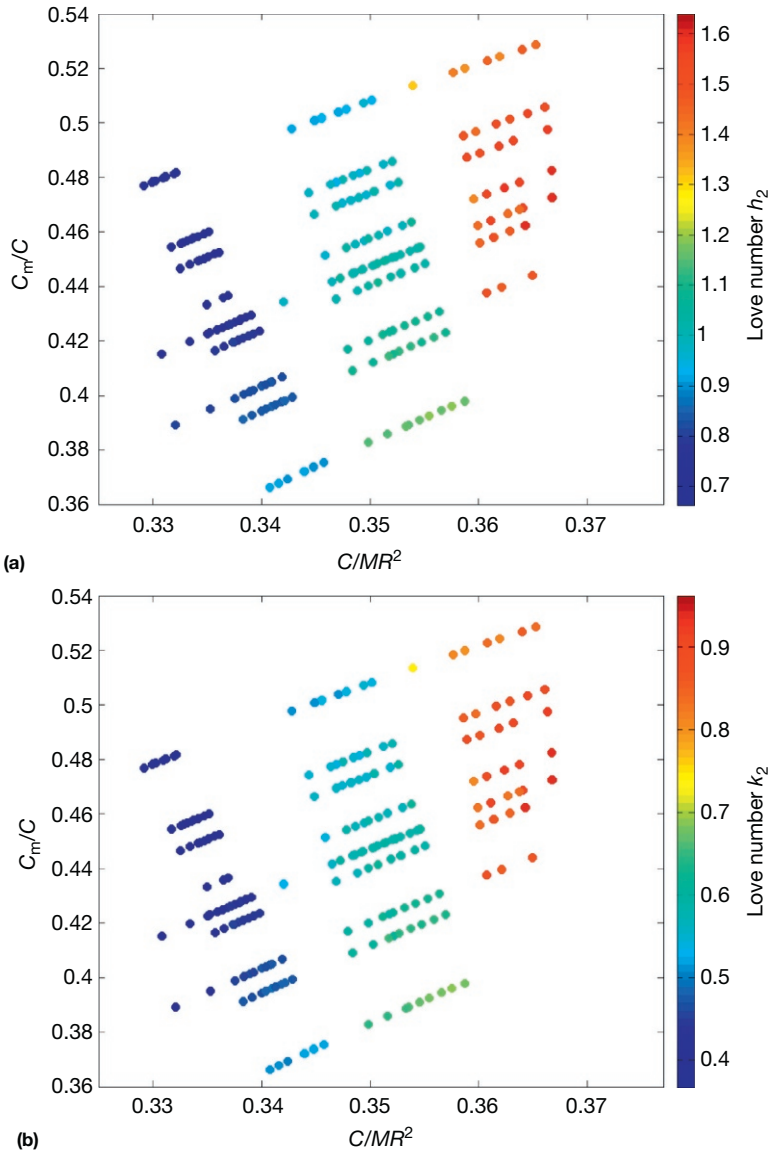


Figure 12 Model determinations of (a) radial displacement Love number h_2 and (b) tidal potential Love number k_2 of Mercury (color-coded) as a function of the axial MoI factor, C/MR^2, and the ratio C_m/C. Reproduced from Steinke T (2012) *Modeling Mercury's Interior Structure and Tidal Deformation*. Bachelor's Thesis (in German), Karlsruhe Institute of Technology, Karlsruhe.

10.02.8.1 General

The Martian surface is characterized by the hemispheric dichotomy between relatively sparsely cratered lowland plains in the north and the heavily cratered southern highland terrain once subjected to the postaccretional heavy bombardment of the inner solar system. The morphology of the boundary between the two hemispheres is dominated by outflow channels and chaotic terrain extending along broad gradual slopes rather than several kilometer high escarpments. Some portions of the dichotomy boundary are composed of fretted terrane and others are sculptured by scarps (Smith and Zuber, 1996). The Tharsis rise is situated close to the near-equatorial boundary between the northern and southern hemispheres and represents a giant volcanic dome established early in the planet's history. Major volcanoes such as Olympus Mons and the Tharsis Montes were emplaced on top of the Tharsis rise. The magmatic evolution of the planet is characterized by a progressive concentration of the volcanic activity to the Tharsis area and, to a lesser extent, to the Elysium area located about 100° to the west in the northern lowland plains. Baratoux et al. (2011) had analyzed Mars Odyssey gamma ray data of major volcanic provinces of Hesperian and Amazonian age and reported compositional trends that can be associated with variable degrees of partial melting of magmatic source regions in the upper Martian mantle. They have inferred a lithospheric growth rate of 17–25 km Gy^{-1} and a time rate of change of potential mantle temperature or secular cooling rate of 30–40 K Gy^{-1} that are in good agreement with estimates of elastic lithosphere thickness from gravity and topography data and numerical models of the thermal and chemical history of Mars, respectively (Baratoux et al., 2011; Grott et al., 2013).

For review articles on the geophysics and evolution of Mars, we refer the reader to Zharkov et al. (1991), Schubert et al. (1992), Zharkov and Gudkova (1997), Spohn et al. (1998, 2001a), Nimmo and Tanaka (2005), and Solomon et al. (2005). Grott et al. (2013) provided an extensive review of Martian igneous processes and their implications for the thermochemical evolution of major geochemical reservoirs (crust, mantle, and core) on Mars.

10.02.8.2 Interior Structure

The global topography and gravitational field of Mars have been determined with high accuracy using laser altimetry and two-way Doppler tracking of the MGS (Konopliv et al., 2006; Lemoine et al., 2001; Smith et al., 1999a,b; Zuber et al., 2000) and Mars Reconnaissance Orbiter (MRO) (Konopliv et al., 2011; Zuber et al., 2007). Doppler radio tracking data of a number of orbiting and landed spacecraft have provided basic information about the planet's mass ($GM = 42828.371901 \pm 0.000074$ km^3 s^{-2}, with G the gravitational constant (Lemoine et al., 2001)), spin-axis precession rate (-7606.1 ± 3.5 mass $year^{-1}$ (Kuchynka et al., 2014)), degree-2 tidal potential Love number k_2 (Konopliv et al., 2006, 2011; Marty et al., 2009; Yoder et al., 2003), and static and seasonal gravitational field. General improvements of the global gravity field solutions have been achieved by combining MGS and MRO tracking data with Mars Odyssey tracking data and surface tracking data from the Pathfinder and Viking 1 landers; the latter permits improvement of the parameters describing the orientation of the planet's rotational axis (Konopliv et al., 2006, 2011).

The rotational flattening of Mars results in a difference between the polar (north–south average) and equatorial radii of about 20 km (Seidelmann et al., 2002, 2005, 2007). The variation of topography with respect to the Martian geoid is about 30 km, representing the largest dynamic range of any terrestrial planet. The hemispheric dichotomy of Mars is related to an offset of about 3 km between the planet's center of mass and center of figure along the polar axis. The Tharsis bulge causes an additional offset of about 1.4 km along an equatorial axis in the direction of Tharsis. The topographic data show that the Tharsis rise consists of two broad rises, a larger, nearly circular southern rise superposed on the highlands that contains the Tharsis volcanoes and a smaller northern rise superposed on the lowlands that contains the shield volcano Alba Patera (Smith et al., 1999a).

Early attempts at modeling the deep interior structure of Mars suffered from poorly known values of its radius and MoI. Improved measurements of the planet's mass M, radius R, gravitational potential, and rotation rate by the Mariner, Viking, and Pathfinder spacecraft provided geodetic constraints required for models of the interior structure. The polar MoI of Mars has been derived from a combined analysis of low-degree gravitational field data and spin-axis precession estimates from MGS tracking and Mars Pathfinder and Viking Lander range and Doppler data. The reanalysis of the entire data set resulted in improved values of $C/MR_e^2 = 0.3650 \pm 0.0012$ (Yoder et al., 2003) and, with the collection of further spacecraft tracking and rotational data, $C/MR_e^2 = 0.3655 \pm 0.0008$ (Konopliv et al., 2006) and $C/MR_e^2 = 0.3644 \pm 0.0005$ (Konopliv et al., 2011), where R_e is Mars' equatorial radius fixed at 3396 km. These values are consistent with the previously accepted value of $C/MR^2 = 0.3662 \pm 0.0017$ (Folkner et al., 1997) that was referred to a mean planet radius R of 3390 km and thus corresponds to $C/MR_e^2 = 0.3649 \pm 0.0017$ if a correction factor $(R/R_e)^2$ is applied. Furthermore, the improved values of the MoI factor of Mars are consistent with the model of a mostly hydrostatic planet with a nonhydrostatic contribution to the MoI factor entirely related to the axisymmetric distribution of topographic loads about Tharsis (Kaula, 1979; Reasenberg, 1977).

Using the improved values of the MoI factor of Mars and taking into account the planet's gravitational oblateness and minor contributions due to the Tharsis rise suggest a stronger concentration of mass toward the center than previously thought, with consequences for the planet's bulk chemistry and interior structure (Konopliv et al., 2011; Sohl et al., 2005; Zharkov and Gudkova, 2005; Zharkov et al., 2009). As a consequence of the improved lower MoI factor, the Martian mantle may be less dense, about several tens of kg m^{-3}, with a smaller iron content than previously thought if crust thickness and core size are specified. It further implies that the Martian crust is several tens of kilometers thicker than previously thought if crust and mantle densities and core size are given. Finally, it suggests several tens of kilometers larger core radii if other parameters like core density, crust density, and crust thickness are fixed (**Figure 13**). If the crust thickness increases, dense mantle material will be replaced by less dense crust material, thereby reducing the planet's MoI factor. The mass deficit that

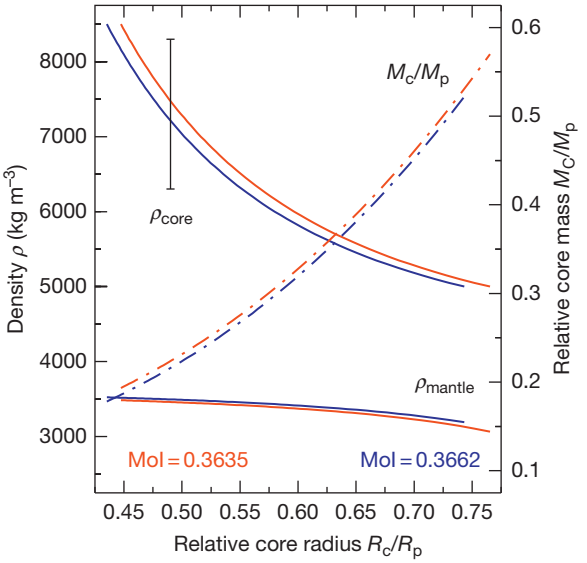


Figure 13 Two-layer model determinations of mantle and core density and core mass fraction M_c/M_p of Mars versus relative core radius R_c/R_p based on an MoI factor of 0.3635 (red) and 0.3662 (blue), respectively. Martian core densities are thought to range between those of γ-iron and iron sulfide as indicated by the error bar. Adapted from Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008, http://dx.doi.org/10.1029/2005JE002520, with permission American Geophysical Union; Copyright 2005.

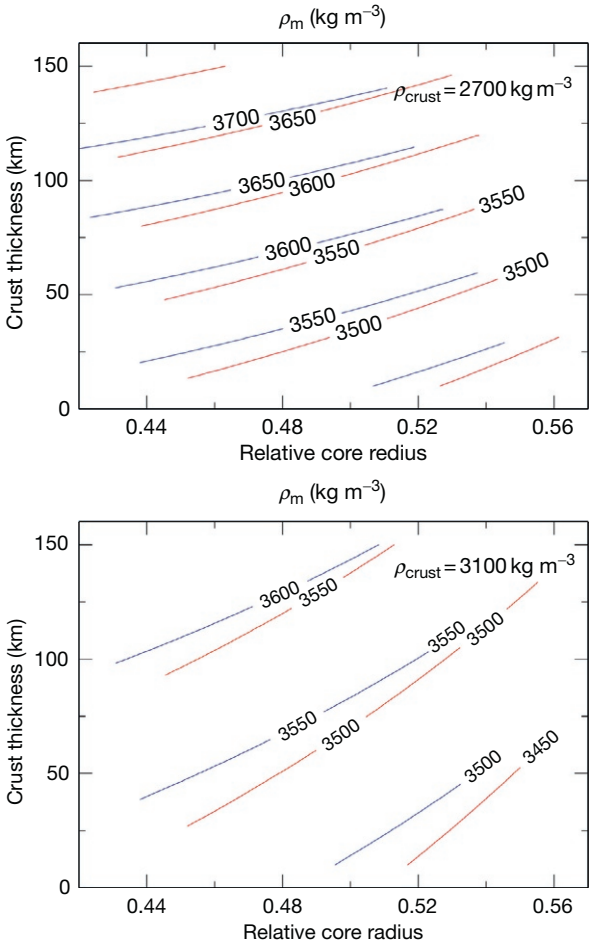


Figure 14 Contours of mantle density ρ_m as a function of relative core radius R_c/R_p and crust thickness based on an MoI factor of 0.3635 (red) and 0.3662 (blue), respectively. Mean crustal density is fixed at (top) 2700 and (bottom) 3100 kg m^{-3}. Adapted from Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008, http://dx.doi.org/10.1029/2005JE002520, with permission American Geophysical Union; Copyright 2005.

arises cannot be compensated simply by increasing the core density, however, since this would even further reduce the MoI factor for a given core size. Therefore, to account for both the planet's mean density and the mean MoI factor at constant core size, the mantle density will be required to increase and core density simultaneously to decrease. In terms of composition, the presence of a thicker crust requires the silicate mantle to be more enriched in iron with an increasing amount of light alloying elements such as sulfur concentrated in the core. Additionally, the contribution of a thicker crust to the planet's bulk composition will become more pronounced (**Figure 14**). However, one should be cautious against changes in the mantle iron concentration that would affect the MoI factor by shifting the locations of pressure-induced mineral phase transitions to other depths (Mocquet et al., 1996, 2011).

Together with the planet's mean density and axial MoI, the solar tidal potential Love number k_2 imposes additional constraints on the size and physical state of the Martian core (Rivoldini et al., 2011; Sohl et al., 2005; Verhoeven et al., 2005; Zharkov and Gudkova, 2005; Zharkov et al., 2009). The determination of $k_2 = 0.153 \pm 0.017$, based on the analysis of 3 years of MGS radio tracking data (Yoder et al., 2003), or $k_2 = 0.152 \pm 0.009$, if combined with additional years of Odyssey tracking data (Konopliv et al., 2006), indicates that the planet's interior is still sufficiently hot that at least the outer part of the Martian core is liquid. Based on a joint analysis of the longest available data set of both missions, Marty et al. (2009) reported values $0.11 < k_2 < 0.13$ for different data subsets. Their preferred solution $k_2 = 0.120 \pm 0.003$ is still consistent with the existence of a dense lower silicate mantle surrounding a liquid metallic core. A larger value $k_2 = 0.164 \pm 0.009$ results from 2 years of MRO radio tracking data, augmented by additional MGS and Mars Odyssey tracking and after correction of the atmospheric tide (Konopliv et al., 2011). The determination of k_2 based on gravitational field observations is largely dependent on data selection and bias effects. This could be a plausible reason for the notable difference between the k_2 estimates by Yoder et al. (2003) and Konopliv et al. (2006, 2011) as compared to that of Marty et al. (2009).

The tidal forcing of Mars allows inference of the planet's bulk rheological properties from the satellites' orbital evolution that is closely connected to the dissipation of tidal energy within the Martian interior (Burns, 1992). Former estimates of Mars' tidal quality factor cover a broad range of $Q = 100 \pm 50$ (Smith and Born, 1976; Yoder, 1982). A small apparent tidal lag angle γ of about 0.7° has been obtained from a combined analysis of Viking and MGS observations of Phobos' orbital position (Bills et al., 2005). The corresponding estimate of

Mars' tidal quality factor is $Q = 1/\tan\gamma \approx 85$. Using orbital position observations of Phobos and Deimos from 1877 to 2005, including MGS and Mars Express observation, Lainey et al. (2007) found $Q = 79.91 \pm 0.69$, provided that $k_2 = 0.152$ and $GM_{Ph} = 0.68 \times 10^6$ m^3 s^{-2} are assumed for Mars' tidal Love number and Phobos' mass, respectively. Zharkov and Gudkova (1997) concluded that the tidal phase lag is not sensitive to the Martian crust and upper mantle, but the dissipation mainly occurs at greater depth in the middle mantle and lower mantle. A small correction on the order of 3 $k_2/Q \sim 0.005$ arises from anelastic softening due to transient or time-dependent creep in the tidal frequency domain. Zharkov and Gudkova (2005) discussed seismic velocity dispersion due to anelasticity. Therefore, the resultant value of the seismic Love number $k_2 = 0.159 \pm 0.009$ should be used as observational constraint on structural models of the Martian interior in addition to the planet's MoI factor. The most recent determinations of k_2 and C/MR_e^2 (Konopliv et al., 2011) are in good agreement with a Martian core radius ranging between 1700 and 1800 km (Konopliv et al., 2006, 2011; Rivoldini et al., 2011; Zharkov et al., 2009).

The first observational estimate of the duration of the Chandler wobble of Mars, a free long-period oscillation of the entire planet, was obtained by Konopliv et al. (2006). In many publications, in which models of the interior structure of Mars were constructed, the predicted model values of the Chandler period were about 200 days. In the observational data, the Chandler term at a period of 210 days is a combination of the free wobble and 1/3 of the annual (229 days) forced wobble that may cause the free wobble signal to shift to longer periods (see **Chapter 10.04** in this volume). Accounting for inelasticity of the planet's interior owing to transient or time-dependent creep, Zharkov and Gudkova (2009) calculated the Chandler wobble period for a set of Martian interior structure models that satisfy the mean MoI factor of $I/MR^2 = 0.3655 \pm 0.0008$ (Konopliv et al., 2006) and corresponding elastic Love numbers $k_2 = 0.145 \pm 0.017$ (Yoder et al., 2003) and $k_2 = 0.148 \pm 0.009$ (Konopliv et al., 2006), respectively. Whereas the minimum Chandler period is $202.8 < T_w < 203.4$ days, its shift to longer periods due to inelasticity δT_w is in the range $0.981 < \delta T_w < 1.42$ days. Combining both estimates, the Chandler wobble period of Mars is found to be in the interval $203.8 < T_w + \delta T_w < 204.8$ days (Zharkov and Gudkova, 2009).

Whereas the gravitational field of the elevated southern hemisphere is relatively featureless and implies a state of near-isostatic compensation, the northern lowland plains reveal a wider range of gravitational anomalies (Smith et al., 1999b). The dichotomy boundary is not clearly resolved on the gravity map, whereas the Tharsis Montes, Olympus Mons, Valles Marineris, and Isidis impact basins are visible as individual gravitational anomalies in the areoid. Major gravity highs are associated with the Tharsis and Elysium volcanoes indicating that they are not isostatically compensated. Large impact basins reveal negative annular anomalies with a central positive anomaly. The global crust and upper mantle structure of Mars has been derived from MGS measurements of gravity and topography. The Bouguer gravity has been interpreted in terms of crustal thickness variations. In these models, the thickness of the southern hemisphere crust decreases progressively from south to north, whereas the northern lowlands are characterized by a more uniform crust thickness (Neumann et al., 2004; Zuber et al., 2000). Elastic thickness models based on gravity and topography data can be used to infer additional information about the planet's rheological structure with important implications for the thermal and volatile evolution of the Martian interior (Guest and Smrekar, 2007; Jakosky and Phillips, 2001).

Plausible Martian crust densities range from 2700 to 3100 kg m^{-3} based on end-member-type compositional models of the Martian crust (Babeyko and Zharkov, 2000; Wieczorek and Zuber, 2004). The lower estimate of about 2700 kg m^{-3} represents an andesitic–basaltic composition obtained from Pathfinder-APXS measurements of soil-free Martian rocks (Brückner et al., 2003). The upper estimate of about 3100 kg m^{-3} (Pauer and Breuer, 2008) represents porous basaltic shergottites (Britt and Consolmagno, 2003) believed to be samples of the Martian crust released during one or several giant impacts (McSween, 1994). These rocks are further believed to represent end-members of the composition of well-mixed Martian soil on a planetary scale (Nimmo and Tanaka, 2005). Using a spectral localization method to analyze volcanic surface loads, Belleguic et al. (2005) had reported considerably higher crustal densities beneath the Elysium rise of 3270 ± 150 kg m^{-3}, whereas the load densities related to the major Martian volcanoes except Alba Patera are best constrained by a value of 3200 ± 100 kg m^{-3}. The latter density is similar to that of Shergotty-type basaltic meteorites thought to have originated from Mars. This implies that the Martian lowland crust is composed of more mafic constituents than the southern highland crust and that the hemispheric difference in elevation may be mainly attributed to a Pratt-like isostatic compensation mechanism (Belleguic et al., 2005). Load densities of ~3350 kg m^{-3} have been obtained for the Tyrrhenus Mons highland volcano provided that low-density pyroclastic deposits make up to 30% of the load volume (Grott and Wieczorek, 2012), suggesting that basaltic lava flows on Mars are compositionally similar and comparable to Martian meteorites. This is supported by high-resolution radio science tracking data of the Tharsis province acquired by the Mars Express (MaRS experiment) and MRO spacecraft. Whereas the major volcanic shields of the Tharsis province, except Ascraeus Mons, were inferred to be mainly composed of high-density lava, the gravitational signature of Olympus Mons could be in agreement with bottom loading in terms of a buoyant mantle plume (Beuthe et al., 2012).

Present estimates of the mean crustal thickness of Mars are entirely based on indirect geophysical studies like local relationships between gravity and topography and/or geochemical arguments and may range between 30 and 80 km (Neumann et al., 2004; Solomon et al., 2005; Wieczorek and Zuber, 2004), although crust thicknesses of up to about 100 km are consistent with global geophysical constraints (Gudkova and Zharkov, 2004; Kavner et al., 2001; Sohl and Spohn, 1997; Sohl et al., 2005). Assuming a Bouguer reduction density of 2900 kg m^{-3} and a crust–mantle density contrast of 600 kg m^{-3}, crust thicknesses vary from about 6 to 102 km, the former of which is located beneath the center of Isidis and the northwestern floor of Hellas Planitia (Neumann et al., 2004). Larger crust thicknesses might be unlikely, however, since ductile flow in the warm lower crust could cause

relaxation of lateral crustal thickness variations (Nimmo and Stevenson, 2001; Zuber, 2001; Zuber et al., 2000). The crust thickness could be further limited by the pressure-induced transition from basalt to eclogite that may occur in the lower Martian crust (Babeyko and Zharkov, 2000). Using the technique of petrophysical modeling (Sobolev and Babeyko, 1994) to construct mineralogical, density, and seismic velocity profiles of the deep Martian crust, Babeyko and Zharkov (2000) first derived the equilibrium mineralogical composition by numerical thermodynamic simulation for a given bulk chemical composition and local P–T conditions, whereas the local density and seismic velocities were calculated in a second step from the mineralogy. Here, it is assumed that the bulk chemical composition of the crust is represented on average by four basaltic Martian meteorite classes (Shergotty, Zagami, EETA 79001, and lithologies A and B).

Vector magnetic field observations, obtained at altitudes between about 100 and 200 km subsequent to orbit insertion of the MGS spacecraft, have been compiled into a global magnetic field map of the Martian crust (Connerney et al., 2005). Although Mars lacks an intrinsic magnetic field at present, the Martian crust acquired intense magnetization in the past, hinting at the existence of a strong internally generated field at that time. Crustal remanent magnetization has been found to exceed that of Earth by more than an order of magnitude and is mainly confined to the most ancient, heavily cratered southern highland terrain (Acuña et al., 1999; Connerney et al., 1999; Quesnel et al., 2007). The bottom of the magnetized layer is constrained by the Curie temperature of its magnetic carriers, whereas the top of the layer is given by the thickness of the uppermost crust that has been demagnetized by impact events. Magnetized source bodies are believed to be situated at depths of less than ~100–90, 90–80, or 55–45 km, provided hematite, magnetite, and pyrrhotite are the magnetic carrier minerals that predominate within the Martian crust (Arkani-Hamed, 2005). Early serpentinization or chemical water–rock reaction at elevated pressure of the Martian crust in the Noachian has been suggested to form enough magnetic carrier minerals such as magnetite (Quesnel et al., 2007). If serpentinization and related crustal density reduction were limited to the southern hemisphere, however, this model could explain not only the apparent concentration of strong magnetic anomalies in the southern highland terrain but also the hemispheric dichotomy of Mars in terms of topography variation and merely subdued gravity changes across the dichotomy boundary (Quesnel et al., 2007, 2009). Early formation of the crustal dichotomy could have resulted in a single-hemisphere dynamo with strong magnetic fields and strongly magnetized crustal units in the southern hemisphere (Stanley et al., 2008).

Linear magnetic features of up to ~2000 km in length are frequently oriented in the east–west direction, some resembling a pattern of band-like features of alternating magnetic polarity (Connerney et al., 2001). The Martian crust may have been magnetized during a time span of only a few hundred million years after planet formation when an active core dynamo driven by thermal convection produced an intense global magnetic field. If crust magnetization occurred later in the planet's evolution, a core dynamo driven by chemical convection associated with inner core growth would be more likely (Connerney et al., 2004). The lack of crustal magnetization in the vicinity of large impact basins (Hellas, Argyre, and Isidis) is commonly attributed to cessation of the core dynamo when these basins formed about 4 Gy during the Noachian period and demagnetized underlying crust layers (Acuña et al., 1999). Alternatively, however, impact basin formation may predate the onset of the Martian core dynamo so that the magnetization of the southern highland terrain was caused by localized heating events followed by rapid cooling below the Curie temperature (Schubert et al., 2000). The timing of the Martian dynamo remains a subject of debate. By analyzing magnetic anomalies in the regions surrounding Tyrrhenus Mons and Syrtis Major, two volcanoes that were active during the late Noachian and Hesperian, Milbury et al. (2012) concluded that the Martian dynamo was active into the Hesperian, perhaps as late as 3.6 Gy. However, Lillis et al. (2013) studied the magnetic anomalies of large Martian craters and claimed that the Martian dynamo ceased operating before the Hellas and Utopia impacts, between 4.0 and 4.1 Gy. The almost complete demagnetization of the Tharsis rise and other volcanic features implies that thermoremanent magnetization on Mars is confined to a relatively thin layer less than a few tens of kilometers thick. In a similar way, a possible former magnetization of the northern lowland crust may have been entirely erased by catastrophic volcanic flooding during the Hesperian period (Head et al., 2002). Connerney et al. (2005) argued that variations in the crustal magnetic field can be associated with major geologic and topographic features, sometimes reminiscent of transform faults in oceanic crust on Earth. It is speculated that arcuate and linear magnetic features, similar to but much larger than those observed near mid-ocean ridges on Earth, could imply that early Noachian crust formation on Mars was accompanied by plate tectonics associated with crustal spreading in the presence of a reversing dynamo (Connerney et al., 2005).

10.02.8.3 Composition

Anders et al. (1971) were among the first who discussed terrestrial planet formation in the presence of planetesimals with different oxidation degrees. This idea was further developed by Ringwood (1977, 1979) and considered in more detail by Wänke (1981, 1991), Dreibus and Wänke (1985, 1989), and Wänke and Dreibus (1994). It appears correct to say that the Dreibus–Wänke (DW) model is as important for Mars as Ringwood's pyrolite model for understanding Earth's interior (Ringwood, 1975).

The DW model is based on considerations of element correlations in SNC meteorites (named after Shergotty, Nakhla, Chassigny, almost certainly representing distinct classes of Martian meteorites) and cosmochemical constraints, suggesting that Mars was formed by the homogeneous accretion of two geochemically distinct components, a highly reduced refractory component and a volatile-rich oxidized one. In this model, the former did not include elements more volatile than sodium or potassium; however, it contained all other elements in the same abundance ratios as in CI chondritic material. The water content of that component was taken to be ~7.3 wt.% (Dreibus and Wänke, 1989). Later, this estimate was increased up to 18–22 wt.% (see also Volume 1 of *Treatise on Geochemistry*). This suggests that the planet captured a substantial amount of water during its formation. Iron and all siderophile

elements are in the metallic state, and even silicon is partly metallic. This material existed mainly in the feeding zone of the growing Earth (Dreibus and Wänke, 1989; Wänke and Dreibus, 1988). In this model, Mars is differentiated into a FeO-rich silicate mantle containing radiogenic heat sources in terrestrial abundances and a sulfur-rich Fe–Ni–FeS core of about 22% of the planet's mass containing about 14 wt.% sulfur with a radius roughly 50% of the surface radius. Following the successful determination of the planet's axial MoI from Mars Pathfinder tracking observations (Folkner et al., 1997), one immediate question was whether the cosmochemical DW model (Dreibus and Wänke, 1985) is consistent with structural models of the Martian interior that satisfy the most recent data on the MoI of Mars (Bertka and Fei, 1997, 1998a; Kavner et al., 2001; Sohl and Spohn, 1997; Yoder and Standish, 1997; Zharkov, 1996; Zharkov and Gudkova, 2000).

Longhi et al. (1992) recasted the mantle composition of Dreibus and Wänke (1985) into a pressure-dependent mineralogy resulting in an upper olivine-rich part, a transition zone composed of silicate spinel, and a lower perovskite-rich layer. The model provides a mean dimensionless MoI factor of 0.353 and is consistent with the geochemical constraint of a bulk Fe/Si ratio of 1.71 representative of the composition of CI carbonaceous chondrites. From models that use identical sets of material parameter values for crust, mantle, and core, as derived from geochemical analyses of SNC meteorites and laboratory studies, and that take into account self-compression and thermal expansion of the Martian interior, a number of authors concluded that it is difficult to reconcile the assumption of a bulk planet CI Fe/Si ratio with the observed value of the axial MoI factor (Bertka and Fei, 1997, 1998a; Sohl and Spohn, 1997; Yoder and Standish, 1997). In contrast, Zharkov and Gudkova (2000) suggested that the observational data basis is not sufficient to assess the validity of the cosmochemical DW model and hence the correctness of the chondritic hypothesis. At present, interior structure models can be constructed that either satisfy or disprove the chondritic hypothesis. Rivoldini et al. (2011) further concluded that geodesy data are not diagnostic of the bulk planet Fe/Si ratio and the mineralogy of the mantle and the crust, even if the thermal state of the Martian mantle is specified.

The geochemical estimate of mantle differentiation and the volatile evolution of Mars are uncertain owing to inherent complexity (Guest and Smrekar, 2007; Jagoutz, 1991), and the formation age and average thickness of the crust are still not well constrained and may vary by a factor of 2 (Wieczorek and Zuber, 2004). Analysis of the SNC meteorites, estimates of the elastic lithosphere thickness, and the dissipation of tidal energy suggest that a few tens of ppm of water are still present in the Martian mantle (Grott et al., 2013). Estimates of the planet's initial bulk water content range from a few ppm (Wänke, 1994) to rheologically significant amounts of ~200 ppm (McCubbin et al., 2010a, 2012a; McSween et al., 2001) for wet mantle conditions, which could result in a melting point reduction of ~100 K (Katz et al., 2003) relative to a dry mantle. If the mantle was wet, magmatic degassing could have supplied a substantial amount of water to the Martian surface early in its history. Using age determinations of zircons and incompatible element abundances of impact melt rocks in a putative ancient southern highland regolith breccia, Humayun et al. (2013) concluded that the magmatic differentiation of the Martian highland crust occurred within the first 100 My after planet formation, a timescale comparable to Earth's and the Moon's, thereby resulting in an average crustal thickness of ~50 km. This suggests to the authors that an equally rapid release of volatiles from the interior is likely and has had important implications for early Martian climate and habitability (Jakosky and Phillips, 2001; Marty and Marti, 2002). It remains unresolved, however, whether the water contents derived from analysis of Shergottites, the most abundant Martian meteorites, correspond to the ancient or present mantle since the formation age of the Shergotty meteorites and the role of magmatic degassing of their source region are still debated (Balta and Mcsween, 2013; Grott et al., 2013).

The Martian core is sufficiently large that even small changes in core size will result in significant changes of core volume. Since the core contributes most of the iron, a significant change in core volume can be expected to result in a notable modification of the bulk planet Fe/Si ratio. Modifications of the Fe/Si ratio due to changes in core volume may be partly compensated, however, by variable mantle iron contents in combination with crust thickness variations because of the significant mass fraction of the planet's silicate portion. The bulk planet Fe/Si ratio is also notably dependent on the thermal state of the core because of the temperature dependence of the equation-of-state parameters of iron and iron sulfide as the main core constituents (Buono and Walker, 2011; Chudinovskikh and Boehler, 2007; Fei et al., 1995; Kavner et al., 2001).

In **Figure 15**, radially symmetrical density models of the Martian interior are compared for variable sulfur contents of

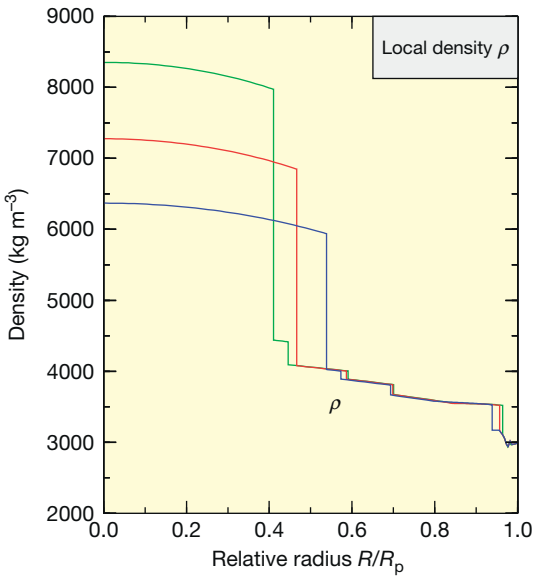


Figure 15 Radial density distribution of the Martian interior as a function of relative core radius R_c/R_p and core composition using the low-temperature crust model of Babeyko and Zharkov (2000). If the core were composed of pure iron (green), it would be small enough to enable the pressure-induced phase transformation from spinel to perovskite. In turn, a pure iron sulfide core (blue) would be too large and mantle pressures therefore too low for the phase transition to occur. The intermediate density profile (red) refers to a Fe–FeS core containing iron and iron sulfide in equal amounts.

the core. The sulfur content of the Martian core is unknown, but geochemical models predict a sulfur-rich core (McSween, 1994). Multianvil-press experiments to examine the effects of both nickel and sulfur on the partitioning of oxygen between liquid Fe metal and ferropericlase suggest that the oxygen content of a more sulfur-rich core could be in the range 2–4 wt.% if it is in equilibrium with the FeO-rich Martian mantle (Tsuno et al., 2011). Furthermore, light constituents such as iron hydride may also be present in the core depending on the amount of hydrogen dissolved in the core alloy when subject to the pressure and temperature conditions possibly prevailing in the Martian core (Gudkova and Zharkov, 2004; Zharkov, 1996; Zharkov and Gudkova, 2000). The way how hydrogen could enter the Martian core is discussed in more detail by Zharkov and Gudkova (2005). If core size and density are specified, more iron-rich alloys can be accommodated with increasing hydrogen content and/or core temperature, thereby increasing the planet's bulk Fe/Si ratio.

10.02.8.4 Mineralogy

Bertka and Fei (1997) obtained the sequence of mineralogical phase assemblages stable in the Martian mantle at elevated temperatures and up to core–mantle boundary pressures from high-pressure and high-temperature experiments using synthetic mineral mixtures according to the mantle composition of Dreibus and Wänke (1985). In **Figure 16**, it is seen that the upper part of the mantle then contains olivine, clinopyroxene, and garnet, whereas orthopyroxene is only present at pressures below 9 GPa. In the mantle transition zone at pressures above 13.5 GPa, β-spinel (wadsleyite) and clinopyroxene are subsequently replaced by γ-spinel (ringwoodite) and majorite until completion at about 17 GPa. A hot lower mantle is found to contain Mg–Fe silicate perovskite, magnesiowüstite, and majorite in the absence of $CaSiO_3$ perovskite (Bertka and Fei, 1997).

Verhoeven et al. (2005) established empirical relationships between the most abundant mantle minerals based on a compilation of chemical compositions and related mineral assemblages of the silicate mantles of Mars and Earth. A perovskite layer at the base of the Martian mantle could only exist if pressures and temperatures in the mantle are sufficiently high for the occurrence of the mineral phase transitions. Therefore, the possible existence of a perovskite lower mantle and its stable phase assemblage in Mars not only depends on the core–mantle boundary pressure, that is, the size and composition of the core, but also is sensitive to the temperature distribution deep inside the planet.

The stability range of such a putative perovskite lower mantle and its consequences for mantle plume dynamics were studied by van Thienen et al. (2006), comparing the model mantle mineralogy from Bertka and Fei (1997) to the model EH45 from Sanloup et al. (1999). Furthermore, uncertainties of a few 100 K in the experimental determinations of the pressure–temperature relation of the perovskite phase transformation and poor knowledge of the planet's thermal state (Breuer and Spohn, 2003) are taken into account. For a nominal core sulfur content of 14 wt.%, van Thienen et al. (2006) found that a perovskite layer sufficiently thick to affect mantle dynamics can be kept for both hot and cold end-member-type mantle temperature profiles (Breuer and Spohn, 2003). Thus,

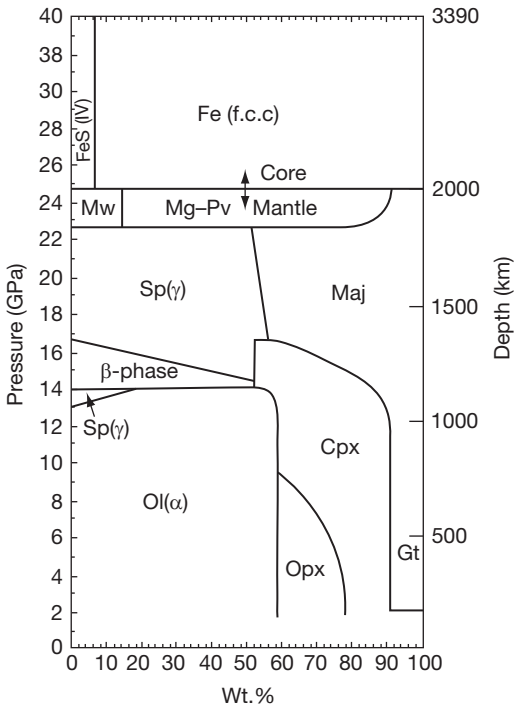


Figure 16 Mineral phase assemblages (in wt.%) of the Martian mantle as a function of pressure based on the chemical compositional model of Dreibus and Wänke (1985). Abbreviations are Ol, olivine; Opx, orthopyroxene; Gt, garnet; Maj, majorite; Sp, spinel; Mw, magnesiowüstite; and Mg–Pv, Mg–Fe silicate perovskite. The location of the core–mantle boundary is from Fei et al. (1995). Adapted from Bertka CM and Fei Y (1997) Mineralogy of the Martian interior up to core–mantle boundary pressures. *Journal Geophysical Research* 102: 5251–5264, with permission American Geophysical Union; Copyright 1997.

it is feasible that at least in the early evolution of the planet, when mantle temperatures were much higher, a thin perovskite layer hovered above the core–mantle boundary. Alternatively, a smaller size of the Martian core in favor of a deep perovskite layer is conceivable if a solid inner core composed of γ-iron would be present and surrounded by a less dense, volatile-rich liquid outer core. Given the present uncertainty on the location of the transition to the perovskite phase, the possible existence of a thin perovskite layer in the lower mantle cannot be excluded even in the presence of a large, fully molten core as suggested by the geodetic data (Rivoldini et al., 2011). A definite answer to this problem has to await the successful deployment and operations of a passive seismological experiment at the Martian surface using broadband seismometers (Gudkova et al., 2014).

If the geochemical requirement of a CI chondrite bulk composition is kept, the mantle density profile obtained from the experimentally determined sequence of mineralogical phase assemblages (Bertka and Fei, 1997) and a core density consistent with a core sulfur content of 14 wt.% would result in a mean MoI factor of 0.354 (Bertka and Fei, 1998a). However, this model that lacks a perovskite layer at the base of the mantle would require a crust thickness of 180–320 km assuming a crustal density of 2700–3000 kg m^{-3}. Since such a crust thickness is considered to be unrealistically large, a CI

chondrite bulk composition of the Martian interior as assumed by Dreibus and Wänke (1985) may be questionable (Bertka and Fei, 1998b). Models using the same mantle density profile and a range of model core compositions but allowing for the MoI factor of 0.3662 as obtained from the Pathfinder measurements generally produce bulk Fe/Si ratios below the CI chondrite value of 1.71. This may suggest that the formation of Mars and the terrestrial planets cannot be explained solely by the accretion of CI carbonaceous chondrite material (Bertka and Fei, 1998b). Using Bayesian inversion methods, Rivoldini et al. (2011) estimated median and probability intervals of geophysical and geochemical key parameters for compositionally distinct interior structure models of Mars. According to these models, the radius of the Martian core is 1794 ± 65 km or ~53% of the planet's radius with a core sulfur content of 16 ± 2 wt.%. As a consequence, these authors claimed that, by adjusting equation-of-state parameters, models of the Martian interior can be constructed, which are in agreement with the mantle composition of Dreibus and Wänke (1985), a core sulfur content of 14 wt.%, a bulk Fe/Si ratio of 1.71, recent determinations of the planet's axial MoI and tidal Love number, and most recent estimates of crust density and thickness.

10.02.8.5 Martian Seismicity

A three-axis short-period seismometer on board the Viking Lander 2 at Utopia Planitia collected more than 600 h of data to explore the seismic environment of Mars (Anderson et al., 1976, 1977). The seismometer was added late in the mission planning and therefore had to be installed on the top of the lander structure. Hence, wind noise contaminated the seismic data. No large seismic events were detected, indicating that Mars is less seismically active than Earth by an estimated three orders of magnitude. A seismic signal from one small quake was possibly recorded. Its tentative analysis suggested that it had a magnitude of 3 and had occurred at a distance of 110 km. Identification of late arrivals implied the presence of a crust with a thickness of 15 km near the landing site. However, it cannot be ruled out that wind gusts mimicked a seismic event.

While the seismicity of Mars is largely unexplored, there are indications that Mars could be quite active. Most morphological features on Mars seem to be only partially isostatically compensated (Zuber et al., 2000). These include the Tharsis rise as well as the Chryse and Amazonis basins. Regions of partial isostatic compensation on Earth are generally the most seismically active. Tectonic stresses implied by a lack of compensation are known to be in the 10 to 100 bar range and stress drops in earthquakes are in the same range. Hence, Mars could be tectonically active, though seismicity may be more localized in the vicinity of preexisting faults. Andrews-Hanna et al. (2008) explained the contractional strain history of Mars as inferred from the ages and locations of Noachian to Early Hesperian strike-slip faults southwest of Tharsis by the interaction of plume-induced contraction during Tharsis formation in the Noachian with uniform contraction due to secular cooling of the entire planet. The best-preserved fault in their mapped set of strike-slip faults has a length of ~200 km and with measured lateral offsets ranging from 5 to 9 km.

The early tectonic evolution of Mars in terms of Tharsis-induced loading stresses superimposed on a uniform contractional stress is thought to be accommodated by quite a number of compressional and extensional surface features such as wrinkle ridges and graben. Based on a global compilation of compressional and extensional faults derived from Mars Orbiter Laser Altimeter (MOLA) shaded relief data, Knapmeyer et al. (2006) had constructed a Martian seismicity model that predicts up to 25 seismic events per year with moment magnitudes greater than four corresponding to seismic moments $\geq 1.26 \times 10^{15}$ N m. Most of these events are expected to occur near the Tharsis rise, but other seismic centers may be located south of Hellas and north of Utopia Planitia (**Figure 17**).

10.02.8.6 Future Exploration

The deployment of seismometers will be central to future missions to Mars that attempt determination of the planet's gross interior structure (Dehant et al., 2000, 2012; Grott et al., 2013; Gudkova et al., 2014; Lognonné et al., 2000). Seismological observations would benefit from greatly improved performance of very broadband seismometers combined with new seismic analysis methods based on the cross correlation of seismic noise recorded by two separate seismic stations on Mars and joint seismic/orbiter detection of meteoroid impacts (Dehant et al., 2012). In a similar fashion, radio science tracking of surface stations would improve considerably by using interferometric methods based on two simultaneous Earth–Mars radio links and multiple landers (Dehant et al., 2012). The seismic and geodetic observations will be augmented by electromagnetic sounding and heat flow measurements. Electromagnetic sounding methods are informative about the electrical conductivity of the Martian mantle, indirectly constraining its composition, temperature, and volatile content (Mocquet et al., 2011). Surface heat flow measurements would provide basic information about the planet's thermal state and the bulk abundance of heat-producing elements (Dehant et al., 2012; Grott et al., 2013), thereby helping to test the chondritic hypothesis. Tide-induced gravity variations raised by Phobos also may permit probing of the deep Martian interior by using a very broadband seismometer (Rivoldini et al., 2009; Van Hoolst et al., 2003). Provided free oscillations of Mars are excited by sufficiently large seismic events or atmospheric pressure variations, seismic data acquired by a single broadband instrument alone could be used to infer the radial distributions of density and seismic velocities and the radius of the Martian core (Gudkova et al., 2014).

10.02.9 Venus

Venus, with a radius of 6051.8 km, is only slightly smaller than Earth, but the small size difference may have important consequences for the planet's interior.

10.02.9.1 General

Unlike Earth, Venus has no magnetic field (Donahue and Russell, 1997; Phillips and Russell, 1987; Russell, 1980). This is likely a consequence of the smaller pressure at the center of

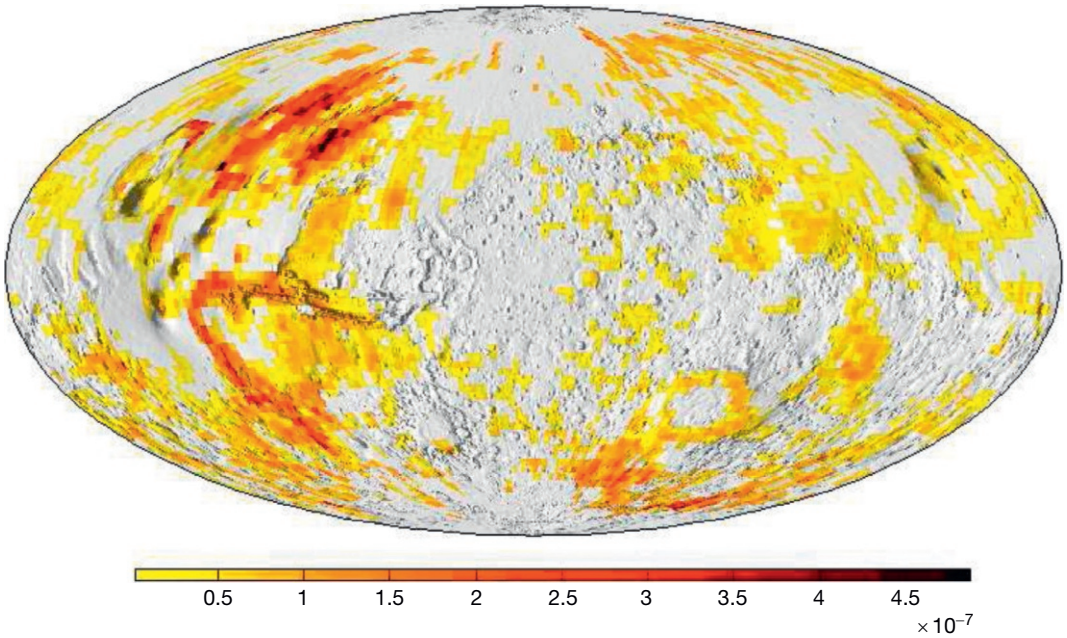


Figure 17 Global distribution of quake probability based on the distribution of extensional and compressional faults on Mars in units of faults per km^{-2}, irrespective of geologic age. Adapted from Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars' seismicity. *Journal of Geophysical Research* 111: E11006, http://dx.doi.org/10.1029/2006JE002708, with permission American Geophysical union; Copyright 2006.

Venus compared with the pressure at the center of the Earth's core. Because of the lower pressure, it is possible that Venus' core has not yet cooled sufficiently to initiate inner core growth, but has cooled enough to prevent the operation of a purely thermally driven dynamo (Stevenson et al., 1983). Venus' lack of a magnetic field could also be due to its lack of plate tectonics, perhaps indicative of a sluggish form of mantle convection that is unable to cool the core efficiently enough to initiate thermal dynamo action (Nimmo and Stevenson, 2000). Still another possibility, though probably unlikely, is that the core of Venus has solidified enough that a dynamo cannot operate in the remaining liquid outer shell (Arkani-Hamed, 1994). Another consequence of Venus' slightly smaller size compared with Earth is that the perovskite–postperovskite phase transition that occurs near the base of the Earth's mantle may not occur in the Venusian mantle. If the core were to contain less light elements than the Earth's core, however, the Venusian mantle could be even deeper than that of the Earth.

10.02.9.2 Interior Structure

The discussion earlier tacitly assumed that the structure of the Venusian interior is similar to that of the Earth, that is, that Venus is basically a three-layer body with a metallic core surrounded by a rocky mantle, which is in turn surrounded by a compositionally distinct rocky crust. The basic structure of Venus is illustrated in Figure 18. However, the structure of the Venusian interior is a matter of some guesswork because we do not know the MoI of the planet, and though spacecraft have visited Venus and landed on its surface, we have not yet seismically probed its interior. Because of the high pressure (~95 bars) and temperature (~737 K) at the surface of Venus (Seiff, 1983), a seismic experiment is probably far into the future. Venus' MoI is also something we will not know anytime soon. Venus lacks a moon to force a precession of the planet's retrograde spin axis, but the torque on the solar-induced tide forces a free precession ranging from 44.1 to 45.6″ $year^{-1}$ and a variation of obliquity, or wobble, estimated at about 1.1″ $year^{-1}$ (Yoder, 1997). Nevertheless, Venus is such a large terrestrial planet and would be heated so thoroughly upon accretion that differentiation into an Earthlike structure is all but inevitable. The extensive basaltic plains on Venus (Basilevsky et al., 1992; Campbell et al., 1997; Surkov, 1983; Weitz and

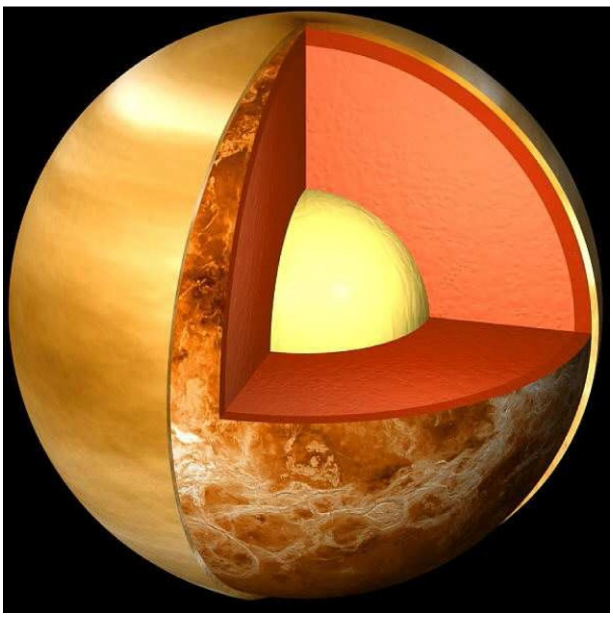


Figure 18 Cutaway view of the interior of Venus. © Calvin J. Hamilton.

Basilevsky, 1993) are evidence that it has differentiated a crust; it is even possible that plateau-like highlands such as Alpha Regio represent compositionally distinct pieces of crust.

Doppler radio tracking data of the Magellan and Pioneer Venus spacecraft have provided basic information about the planet's mass ($GM = 324858.6 \pm 0.014$ km^3 s^{-2}, G is the gravitational constant), gravitational field, and tidal Love number k_2 (Sjogren et al., 1997). The Love number k_2 is determined from the time variations in the gravitational coefficients $C_{2,2}$ and $S_{2,2}$ at the solar period. Konopliv and Yoder (1996) had found $k_2 = 0.295 \pm 0.066$, a value consistent with a liquid core (Yoder, 1997). A 180 degree and order spherical harmonic model of the Venusian gravity is given in Konopliv et al. (1999). Magellan radar altimetry data have been used to produce the 360 degree and order spherical harmonic model of Venus' topography presented in Rappaport et al. (1999) (see **Chapter 10.05** in this volume).

10.02.9.3 Composition

The predominantly basaltic nature of Venus' surface is suggested by the geochemical data obtained by the Venera and Vega landers and the morphology of the widespread volcanic landforms (Grimm and Hess, 1997). The K/U and K/Th ratios of Venusian samples are similar to those obtained for terrestrial volcanics, SNC meteorites, and Martian samples. The variation of the K/U ratios at the seven landing sites is relatively narrow within a factor of three of each other implying that the volatile-refractory element inventory of Venus is comparable to that of Earth and Mars but different from those of the Moon. The major oxide compositions of Venusian rocks are broadly consistent with those of basaltic rocks so that it is reasonable to assume that the crust is largely basaltic (Grimm and Hess, 1997). The gravity and topography data can be analyzed to provide information on the average thickness of the crust; estimates lie in the range 20–50 km (Grimm and Hess, 1997; Nimmo and McKenzie, 1998). The crust is thicker (45–85 km) beneath the plateau highlands (Alpha, Ovda, Thetis, and Tellus Regiones) (Moore and Schubert, 1997) compared with other regions. The basalt–eclogite phase change limits the crustal thickness to about 50 km for moderate conductive geotherms of >5–10 K km^{-1} (Grimm and Hess, 1997). The Venusian gravity and topography data have also been used to infer values for the thickness of the planet's thermal lithosphere and its elastic upper layer. Estimates of lithosphere thickness vary between about 200 and 400 km with thinner lithosphere beneath volcanic highlands (e.g., Atla and Beta Regiones) (Herrick et al., 1989; Kucinskas and Turcotte, 1994; Moore and Schubert, 1997; Phillips, 1994; Simons et al., 1994; Smrekar, 1994). Lithospheric thinning beneath the volcanic rises could have been caused by mantle plumes. There is buoyant sublithospheric mantle beneath the volcanic rises (Moore and Schubert, 1997). The top part of the lithosphere that behaves elastically is only about 30 km thick (McKenzie and Nimmo, 1997). Unlike on Earth, gravity anomalies correlate with high topography on Venus.

10.02.9.4 Tectonism

Pioneer Venus radar, Earth-based radar observations, Venera 15–16 orbital imaging radar, and Magellan radar images have provided views of the surface of Venus unimpeded by the global cloud cover that prevents visual observation. These views, together with the topography and gravity data, reveal the nature of Venusian tectonism and volcanism. On Earth, the global oceanic rift system and the arcuate ocean trenches are the primary surface manifestations of plate tectonics. The almost total absence of these features on Venus has led to the conclusion that active plate tectonics is not occurring on Venus (Kaula, 1994; Kaula and Phillips, 1981). At the present time, Venus is a one-plate planet. Nevertheless, there are tectonic features on Venus that resemble major tectonic features on Earth. Beta Regio, a volcanic highland, has many of the features of a continental rift on Earth. It has a domal structure with a diameter of about 2000 km and a swell amplitude of about 2 km. It has a well-defined central rift valley with a depth of 1–2 km, and there is some evidence for a three-armed planform (aulacogen). Alta, Eistla, and Bell Regiones have similar rift zone characteristics (Grimm and Phillips, 1992; Senske et al., 1992). Aphrodite Terra with a length of some 1500 km is reminiscent of major continental collision zones on Earth, such as the mountain belt that extends from the Alps to the Himalayas. Ishtar Terra is a region of elevated topography with a horizontal scale of 2000–3000 km. A major feature is Lakshmi Planum, which is an elevated plateau similar to Tibet with a mean elevation of about 4 km.

This plateau is surrounded by linear mountain belts, Akna, Danu, Freyja, and Maxwell Montes, reaching elevations of 10 km, similar in scale and elevation to the Himalayas (Kaula et al., 1997).

10.02.9.5 Dynamics

The impact crater population on the surface of Venus has been used to infer a mean surface age of several hundred to as much as 800 My (Herrick et al., 1997; McKinnon et al., 1997). It has been proposed that the relatively young age of Venus' surface was set in a global volcanic resurfacing event and that relatively little volcanism has occurred since (Basilevsky et al., 1997; Schaber et al., 1992). The resurfacing event could be the means by which Venus expels its heat. One way this could happen is the global foundering of a thick, relatively cold, and heavy lithosphere and its replacement by the relatively hot underlying mantle (Turcotte, 1993). Such events might have occurred episodically throughout Venus' history. Between such events, the lithosphere would thicken but Venus would have no efficient way, like plate tectonics on Earth, to expel its heat. Instead, the heat building up in the interior during the quiescent period would be lost in the mantle overturn when the lithosphere thickened enough to become gravitationally unstable. The initiation of such an event might be evident today on Venus' surface in the form of large coronae. Coronae are quasi-circular features, 100–2600 km in diameter, with raised interiors and elevated rims, often with annular troughs (Stofan et al., 1997). Mckenzie et al. (1992) and Sandwell and Schubert (1992a,b) had argued that the perimeters of several large coronae on Venus, specifically Artemis, Latona, and Eithinoa, resemble terrestrial subduction zones in both planform and topography. Artemis chasma has a radius of curvature similar to that of the South Sandwich subduction zone on the Earth. Sandwell and Schubert (1992a) proposed that the large coronae

are incipient circular subduction zones. The foundering lithosphere is replaced by ascending hot mantle in a manner similar to back-arc spreading on the Earth. A single global resurfacing event for Venus has been challenged by Hauck et al. (1998). They argued that the interpretation of the Venusian crater distribution is nonunique, and they identified some units in the volcanic plains that have a spread in age of about 0.5 Gy.

Observations by Venus Express, a European Space Agency mission in orbit around Venus since April 2006, has provided evidence of geologically recent volcanism. The instruments on Venus Express look mainly at the planet's atmosphere, but radiation emitted from the surface at certain wavelengths in the near-infrared reaches the spacecraft through the thick carbon dioxide atmosphere and provides information about surface properties. In this way, the visible and infrared thermal imaging spectrometer on Venus Express mapped the distribution of thermal emissivity over Venus surface (coverage was not complete). Anomalously high values of emissivity were observed at the three hotspots Imdr, Themis, and Dione Regiones (areas analogous to Hawaii, with volcanism, broad topographic rises, and large positive gravity anomalies suggesting mantle plumes at depth). The emissivity anomalies are interpreted to be associated with geologically young lava flows that have experienced relatively little surface weathering (Smrekar et al., 2010). These authors estimate the flows to be younger than 2.5 million years and probably much younger, about 250 000 years or less, indicating that Venus is actively resurfacing.

Because Venus lacks plate tectonics, convection in its mantle is different from the style of convection in the Earth's mantle. Venusian mantle convection occurs in the sluggish or stagnant lid regime, that is, it is confined below the lithosphere or nearly rigid lid (Schubert et al., 1997). This form of convection is less efficient at transporting heat than is the plate tectonic regime with consequent implications for the thermal history of the planet and the dynamics of its core, as noted earlier. Mantle convection in Venus may be unable to establish a near-equilibrium with its internal heat sources resulting in the episodic overturning of its mantle and global resurfacing. Enhanced core cooling would occur during such an event with the possible initiation of a transient dynamo. Heat from the core would be carried away by mantle plumes that could form volcanic rises similar to Atla and Beta Regiones.

10.02.10 Solid Exoplanets

Solid exoplanets with masses of up to about ten Earth masses are thought to have similar interior structures and bulk compositions as the terrestrial bodies in the solar system. Their interiors are thought to be composed of rock-forming elements and metals such as iron, the latter evenly distributed or concentrated in central cores (Elkins-Tanton and Seager, 2008). Unlike for the solar system inner planets, there are even fewer constraints than unknowns in the case of solid extrasolar planets, and even basic interior structure models suffer from inherent nonuniqueness. To address these degeneracies, usually plausible assumptions about their composition and its depth dependence are made (see **Chapter 10.21** in this volume).

Numerical models of planetary interiors using laboratory data on material properties aim at improving the general understanding of their origins, internal evolutions, and present thermal states. In the case of the rocky planets within the solar system, the resultant radial profiles are required to be consistent with geophysical observations and cosmochemical evidence for the compositions of crust, mantle, and core (Mocquet et al., 2011). For rocky exoplanets, the numerical models have to be consistent with the observed planetary masses and radii. Calculated models have been used to derive mass–radius relationships for exoplanets assuming a range of different mineralogical compositions to gain insight into the interior structure and possible bulk compositions of these planets (Fortney et al., 2007; Grasset et al., 2008; Seager et al., 2007; Sotin et al., 2007; Swift et al., 2012; Valencia et al., 2006, 2007; Wagner et al., 2011; Zeng and Sasselov, 2013). Principal uncertainties mainly arise from the extrapolation of an equation of state to high pressures owing to the lack of reliable experimental data in the warm dense matter regime in the pressure range of 200 GPa–10 TPa, whereas the surface temperature and internal thermal state of a massive rocky exoplanet are less important for its radial density distribution (Seager et al., 2007). Nevertheless, the latter are expected to have severe consequences for rheological properties and geodynamic processes. Furthermore, scaling laws for key physical and chemical properties have been obtained (Stamenković et al., 2012; Valencia and O'Connell, 2009; Wagner et al., 2012), which are essential for a better understanding of global planetary processes controlling the general evolution of a planetary body and its astrobiological potential to be life-sustaining.

Figure 19 shows modeled mass–radius relationships in comparison with the relatively large (1–σ) error bars obtained for low-mass planets to date (Wagner, 2014). For the smallest planets, radii appear to be better constrained than masses. These planets are usually detected by space missions (CoRoT and Kepler) providing photometrically accurate light curves, hence radii, but the target objects are often too faint to allow for an accurate mass determination. The knowledge of mean planet density is foremost dependent on the quality of the stellar mass and radius determinations that feeds into the determinations of planetary mass and radius. One of the main goals of future missions is thereby to provide highly precise and accurate measurements of the planet host stars' characteristics, in particular their radii, masses, and ages (Rauer et al., 2014).

10.02.11 Summary and Outlook

The terrestrial planets Mercury, Venus, Earth, and Mars have low masses, small radii, and large densities in comparison with the giant planets in the outer solar system. This is also true for terrestrial-type bodies like the Moon and some of the outer planet satellites, and it provides important clues on their bulk compositions. Rotational, gravitational, and magnetic field observations indicate that the interiors of these bodies are strongly differentiated and subdivided like that of the Earth into iron-rich cores, silicate mantles, and rocky crusts derived from partial mantle melts. Isotope data reveal that the cores and the primary crusts formed early and rapidly. Geodetic observations of the rotational state and/or tidal response suggest that the interiors are warm enough to maintain liquid outer core shells or entirely liquid cores. For Mars, Venus, and Earth, mantle pressures are sufficient to permit mineral

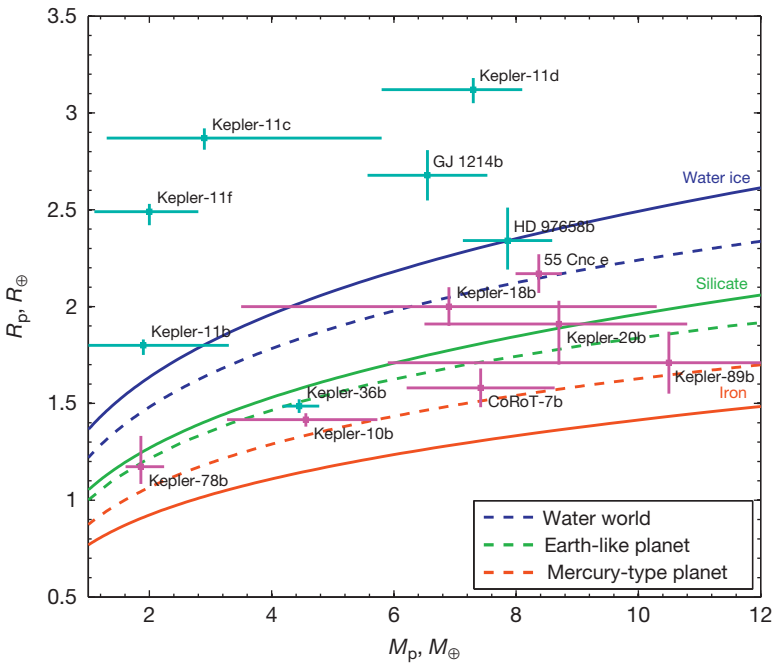


Figure 19 Mass–radius diagram for planets with different bulk compositions compared to currently known low-mass exoplanets in Earth units. Equilibrium surface temperatures are divided into two domains from 500 to 1000 K (cyan blue) and 1000–2000 K (magenta). While the solid curves denote homogeneous, self-compressible solid spheres of water ice, silicate rock, and iron, respectively, the dashed curves exhibit differentiated models of intermediate bulk compositions. Reproduced from Wagner FW (2014) *The Physical State of Rocky Exoplanet Interiors*. Dissertation, Westphalian Wilhelms-University, Munster.

phase transformations from olivine and pyroxene assemblages to spinel or even perovskite and postperovskite phases. Since the phase transition depths also depend on the ambient temperature and the iron content of the mantle rocks, future seismological observations complemented by electromagnetic induction data and heat flow measurements have the potential to provide additional information on the thermal states and compositional differences of the terrestrial planets. Single-plate planets, the Moon, Mercury, Mars, and Venus, are believed to be cooling by lithospheric thickening, while the deep interior remains relatively warm. It is likely, therefore, that due to the progressive cooling of the planet's outer portion, thermoelastic stresses will be occasionally released at preexisting faults, thereby causing local seismic activity at a level detectable by seismometers. The discovery of rocky exoplanets relies on current detection limits of ground-based observational methods. Structural models of solid exoplanet interiors can be constructed by using equations of state for the radial density distribution, which are compliant with the thermodynamics of the high-pressure limit. The modeling results imply that mass–radius relationships are robust and can be used to classify low-mass exoplanets such as CoRoT-7b and Kepler-10b in terms of their bulk compositions.

Acknowledgments

We thank Matthias Grott, Tamara Gudkova, Martin Knapmeyer, Jürgen Oberst, Heike Rauer, Attilio Rivoldini, Alexander Stark, Teresa Steinke, Diana Valencia, and Frank W. Wagner for their helpful discussions. We gratefully acknowledge constructive criticisms from an anonymous reviewer, Raymond Jeanloz, Mark Wieczorek, Tim Van Hoolst, Vladimir Zharkov, and Tilman Spohn. Some figures were produced by using the Generic Mapping Tools (GMT) of Wessel and Smith (1991).

References

Acuña MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.

Acuña MH, Connerney JEP, Wasilewski P, et al. (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars Global Surveyor mission. *Science* 279: 1676–1680.

Akimoto S and Fudisawa H (1968) Olivine-spinel solid solution equilibrium in the system Mg_2SiO_4-Fe_2SiO_4. *Journal of Geophysical Research* 73: 1467–1479.

Anders ER, Ganapathy R, Keays RR, et al. (1971) Volatile and siderophile elements in lunar rocks: Comparison with terrestrial and meteoritic basalts. *Proceedings of the Lunar Science Conference* 2: 1021–1036.

Anderson OL (1984) A universal thermal equation-of-state. *Journal of Geodynamics* 1: 185–214.

Anderson OL and Baumgardner JR (1980) Equations of state in planet interiors. In: *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 1999–2014. Lunar and Planetary Science Institute.

Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The mass, gravity field, and ephemeris of Mercury. *Icarus* 71: 337–349.

Anderson OL, Isaak DG, and Oda H (1992) High-temperature elastic constant data on minerals relevant to geophysics. *Reviews of Geophysics* 30: 57–90.

Anderson BJ, Johnson CL, Korth H, et al. (2011) The global magnetic field of Mercury from MESSENGER orbital observations. *Science* 333: 1859–1862.

Anderson BJ, Johnson CL, Korth H, et al. (2012) Low-degree structure in Mercury's planetary field. *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2012JE004159.

Anderson JD, Jurgens RF, Lau EL, Slade MA III, and Schubert G (1996) Shape and orientation of Mercury from radar ranging data. *Icarus* 124: 690–697.
Anderson DL, Miller WF, Latham GV, et al. (1976) The Viking seismic experiment. *Science* 194: 1317–1321.
Anderson DL, Miller WF, Latham GV, et al. (1977) Seismology on Mars. *Journal of Geophysical Research* 82: 4524–4546.
Anderson JD and Schubert G (2007) Saturn's satellite Rhea is a homogeneous mix of rock and ice. *Geophysical Research Letters* 34: L02202: http://dx.doi.org/10.1029/2006GL028100.
Andrews-Hanna JC, Asmar SW, Head JW III, et al. (2013) Ancient igneous intrusions and early expansion of the Moon revealed by GRAIL gravity gradiometry. *Science* 339: 675–678.
Andrews-Hanna JC, Zuber MT, and Hauck SA II (2008) Strike-slip faults on Mars: Observations and implications for global tectonics and geodynamics. *Journal of Geophysical Research* 113: E08002: http://dx.doi.org/10.1029/2007JE002980.
Araki H, Tazawa S, Noda H, et al. (2009) Lunar global shape and polar topography derived from Kaguya-LALT laser altimetry. *Science* 323: 897–900.
Arkani-Hamed J (1994) On the thermal evolution of Venus. *Journal of Geophysical Research* 99: 2019–2033.
Arkani-Hamed J (2005) Magnetic crust of Mars. *Journal of Geophysical Research* 110: E08005: http://dx.doi.org/10.1029/2004JE002397.
Babeyko AY and Zharkov VN (2000) Martian crust: A modeling approach. *Physics of the Earth and Planetary Interiors* 117: 421–435.
Badding JV, Hemley RJ, and Mao HK (1991) High-pressure chemistry of hydrogen in metals: In situ study of iron hydride. *Science* 253: 421–424.
Balog PS, Secco RA, Rubie DC, and Frost DJ (2003) Equation of state of liquid Fe-10 wt% S: Implications for the metallic cores of planetary bodies. *Journal of Geophysical Research* 108(B2): 2124. http://dx.doi.org/10.1029/2001JB001646.
Balogh A and Giampieri G (2002) Mercury: The planet and its orbit. *Reports on Progress in Physics* 65: 529–560.
Balta JB and Mcsween HY Jr. (2013) Water and the composition of Martian magmas. *Geology* 41: 1115–1118.
Banks R (1969) Geomagnetic variations and the electrical conductivity of the upper mantle. *Geophysical Journal of the Royal Astronomical Society* 17: 457–487.
Baratoux D, Toplis MJ, Monnereau M, and Gasnault O (2011) Thermal history of Mars inferred from orbital geochemistry of volcanic provinces. *Nature* 472: 338–341.
Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1047–1084. Tucson, AZ: University of Arizona Press.
Basilevsky AT, Nikolaeva OV, and Weitz CM (1992) Geology of the Venera-8 landing site region from Magellan data – Morphological and geochemical considerations. *Journal of Geophysical Research* 97(E10): 16,315–16,335.
Bass J (2004) *Current and Future Research Directions in High-Pressure Mineral Physics*. Stony Brook, NY: Consortium for Materials Properties Research in Earth Sciences (COMPRES).
Belleguic V, Lognonne P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005. http://dx.doi.org/10.1029/2005JE002437.
Bertka CM and Fei Y (1997) Mineralogy of the Martian interior up to core-mantle boundary pressures. *Journal of Geophysical Research* 102: 5251–5264.
Bertka CM and Fei Y (1998a) Density profile of an SNC model Martian interior and the moment of inertia factor of Mars. *Earth and Planetary Science Letters* 157: 79–88.
Bertka CM and Fei Y (1998b) Implications of Mars Pathfinder data for the accretion history of the terrestrial planets. *Science* 281: 1838–1840.
Beuthe M, Maistre SL, Rosenblatt P, Pätzold M, and Dehant V (2012) Density and lithospheric thickness of the Tharsis Province from MEX MaRS and MRO gravity data. *Journal of Geophysical Research* 117: E04002. http://dx.doi.org/10.1029/2011JE003976.
Bills BG and Ferrari AJ (1977) A lunar density model consistent with topographic, gravitational, librational, and seismic data. *Journal of Geophysical Research* 82: 1306–1314.
Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004. http://dx.doi.org/10.1029/2004JE002376.
Bina CR and Helffrich GR (1992) Calculation of elastic properties from thermodynamic equation of state principles. *Annual Review of Earth and Planetary Sciences* 20: 527–552.
Binder AB (1998) Lunar Prospector: Overview. *Science* 281: 1475–1476.
Birch F (1952) Elasticity and constitution of the Earth's interior. *Journal of Geophysical Research* 57: 227–286.
Blewett DT, Chabot NL, Denevi BW, et al. (2011) Hollows on Mercury: MESSENGER evidence for geologically recent volatile-related activity. *Science* 333: 1856–1859.
Blewett DT, Lucey PG, Hawke BR, Ling GG, and Robinson MS (1997) A comparison of Mercurian reflectance and spectral quantities with those of the Moon. *Icarus* 129: 217–231.
Boehler R (1992) Melting of the Fe-FeO and the Fe-FeS systems at high pressure: Constraints on core temperatures. *Earth and Planetary Science Letters* 111: 217–227.
Boehler R (1996a) Fe-FeS eutectic temperatures to 620 kbar. *Physics of the Earth and Planetary Interiors* 96: 181–186.
Boehler R (1996b) Melting of mantle and core materials at very high pressures. *Philosophical Transactions of the Royal Society of London A* 354: 1265–1278.
Boehler R (1996c) Melting temperature of the Earth's mantle and core: Earth's thermal structure. *Annual Review of Earth and Planetary Sciences* 24: 15–40.
Boyce JW, Liu Y, Rossman GR, et al. (2010) Lunar apatite with terrestrial volatile abundances. *Nature* 466: 466–469.
Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics on Mars: Evidence for magnetic field history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072. http://dx.doi.org/10.1029/2002JE001999.
Britt DT and Consolmagno GJ (2003) Stony meteorite porosities and densities: A review of the data through 2001. *Meteoritics and Planetary Science* 38: 1161–1180.
Brückner J, Dreibus G, Rieder G, and Wanke H (2003) Refined data of Alpha-Proton-X-ray spectrometer analyses of soils and rocks at the Mars Pathfinder site: Implications for surface chemistry. *Journal of Geophysical Research* 108(E12): 8094. http://dx.doi.org/10.1029/2003JE002060.
Bullen KE (1947) *Introduction to the Theory of Seismology*. Cambridge: Cambridge University Press.
Buono AS and Walker D (2011) The Fe-rich liquidus in the Fe-FeS system from 1 bar to 10 GPa. *Geochimica et Cosmochimica Acta* 75: 2072–2087.
Burns JA (1976) Consequences of the tidal slowing of Mercury. *Icarus* 28: 453–458.
Burns JA (1992) Contradictory clues as to the origin of the Martian moons. In: Kieffer HH, Jakosky M, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1283–1301. Tucson, AZ: University of Arizona Press.
Bussey DBJ and Spudis PD (2000) Compositional studies of the Orientale, Humorum, Nectaris, and Crisium lunar basins. *Journal of Geophysical Research* 105: 4235–4244.
Cameron AGW, Fegley B, Benz W, and Slattery WL (1988) The strange density of Mercury: Theoretical considerations. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.
Campbell BA, Arvidson RE, Shepard MK, and Brackett RA (1997) Remote sensing of surface processes. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 503–526. Tucson, AZ: University of Arizona Press.
Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.
Chapman S and Price AT (1930) The electric and magnetic state of the interior of the earth, as inferred from terrestrial magnetic variations. *Philosophical Transactions of the Royal Society of London A* 229: 0427–0460.
Chase SC, Miner ED, Morrison D, Münch G, and Neugebauer G (1976) Mariner 10 infrared radiometer results: Temperatures and thermal properties of the surface of Mercury. *Icarus* 28: 565–578.
Chenet H, Lognonne P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from the Apollo seismic data set. *Earth and Planetary Science Letters* 243: 1–14.
Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.
Christensen UR and Wicht J (2006) Models of magnetic field generation in partly stable planetary cores: Applications to Mercury and Saturn. *Icarus* 196: 16–34.
Chudinovskikh L and Boehler R (2007) Eutectic melting in the system Fe-S to 44 GPa. *Earth and Planetary Science Letters* 257: 97–103.
Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111: 1–32.
Connerney JEP, Acuña MH, Ness NF, et al. (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences of the United States of America* 102: 14,970–14,975.
Connerney JEP, Acuña MH, Wasilewski PJ, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
Connerney JEP, Acuña MH, Wasilewski PJ, et al. (2001) The global magnetic field of Mars and implications for crustal evolution. *Geophysical Research Letters* 28: 4015–4018.
Constable S and Constable C (2004) Observing geomagnetic induction in magnetic satellite measurements and associated implications for mantle conductivity.

Geochemistry Geophysics Geosystems 5. http://dx.doi.org/10.1029/2003GC000634.
Dehant V, Banerdt B, Lognonne P, et al. (2012) Future Mars geophysical observatories for understanding its internal structure, rotation, and evolution. *Planetary and Space Science* 68: 123–145.
Dehant V, Defraigne P, and Van Hoolst T (2000) Computation of Mars' transfer function for nutations, tides, and surface loading. *Physics of the Earth and Planetary Interiors* 117: 385–395.
Dehant V, Folkner W, Renotte E, et al. (2009) Lander radioscience for obtaining the rotation and orientation of Mars. *Planetary and Space Science* 57: 1050–1067.
Dehant V, Maistre SL, Rivoldini A, et al. (2011) Revealing Mars' deep interior: Future geodesy missions using radio links between landers, orbiters, and the Earth. *Planetary and Space Science* 59: 1069–1081.
Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo program. *Science* 265: 482–490.
Donahue TM and Russell CT (1997) The Venus atmosphere and ionosphere and their interaction with the solar wind: An overview. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 3–31. Tucson, AZ: University of Arizona Press.
Drake MJ (2001) The eucrite/Vesta story. *Meteoritics* 36: 501–513.
Dreibus G and Wänke H (1985) Mars: A volatile rich planet. *Meteoritics* 20: 367–382.
Dreibus G and Wänke H (1989) Supply and loss of volatile constituents during the accretion of terrestrial planets. In: Atreya SK, Pollack JB, and Matthews MS (eds.) *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 268–288. Tucson, AZ: University of Arizona Press.
Dwyer CA, Stevenson DJ, and Nimmo F (2011) A long-lived lunar dynamo driven by continuous mechanical stirring. *Nature* 479: 212–214.
Elkins-Tanton LT, Chatterjee N, and Grove TL (2003a) Experimental and petrologic constraints on lunar differentiation from the Apollo 15 green picritic glasses. *Meteoritics and Planetary Science* 38: 515–527.
Elkins-Tanton LT, Parmentier EM, and Hess PC (2003b) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38: 1753–1771.
Elkins-Tanton LT and Seager S (2008) Ranges of atmospheric mass and composition of super-Earth exoplanets. *Astrophysical Journal* 685: 1237–1246.
Elphic RC and Russell CT (1978) On the apparent source depth of planetary magnetic fields. *Geophysical Research Letters* 5: 211–214.
Esposito PB, Banerdt WB, Lindal GF, et al. (1992) Gravity and topography. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 209–248. Tucson, AZ: University of Arizona Press.
Fassett CI, Kadish SJ, Head JW III, Solomon SC, and Strom RG (2011) The global population of large craters on Mercury and comparison with the Moon. *Geophysical Research Letters* 38: L10202: http://dx.doi.org/10.1029/2011GL047294.
Fei Y, Bertka CM, and Finger LW (1997) High-pressure iron-sulfur compound, Fe_3S_2, and melting relations in the Fe-FeS system. *Science* 268: 1621–1623.
Fei Y, Prewitt CT, Mao HK, and Bertka CM (1995) Structure and density of FeS at high pressure and high temperature and the internal structure of Mars. *Science* 268: 1892–1894.
Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.
Fortney JJ, Marley MS, and Barnes JW (2007) Planetary radii across five orders of magnitude in mass and stellar insolation: Application to transits. *Astrophysical Journal* 659: 1661–1672.
Frey HV, Roark JH, Shockey KM, Frey EL, and Sakimoto SEH (2002) Ancient lowlands on Mars. *Geophysical Research Letters* 29(10). http://dx.doi.org/10.1029/2001GL013832.
Fu RR, Weiss BP, Shuster DL, et al. (2012) An ancient core dynamo in asteroid Vesta. *Science* 338: 238–241.
Gagnepain-Beyneix J, Lognonné P, Chenet H, Lombardi D, and Spohn T (2006) A seismic model of the lunar mantle and constraints on temperature and mineralogy. *Physics of the Earth and Planetary Interiors* 159: 140–166.
Garcia RF, Gagnepain-Beyneix J, Chevrot S, and Lognonné P (2011) Very preliminary reference Moon model. *Physics of the Earth and Planetary Interiors* 188: 96–113.
Ghosh A and Mcsween HJ Jr. (1998) A thermal model for the differentiation of asteroid 4 Vesta, based on radiogenic heating. *Icarus* 134: 187–206.
Glassmeier K-H, Auster U, and Motschmann U (2007a) A feedback dynamo generating Mercury's magnetic field. *Geophysical Research Letters* 34: L22201. http://dx.doi.org/10.1029/2007GL031662.
Glassmeier K-H, Grosser J, Auster U, Constantinescu D, Narita Y, and Stellmach S (2007b) Electromagnetic induction effects and dynamo action in the Hermean system. *Space Science Reviews* 132(2–4): 511–527.
Goins NR, Dainty AM, and Toksöz MN (1981) Lunar seismology: The internal structure of the Moon. *Journal of Geophysical Research* 86: 5061–5074.
Gómez-Pérez N and Solomon SC (2010) Mercury's weak magnetic field: A result of magnetospheric feedback? *Geophysical Research Letters* 37: L20204. http://dx.doi.org/10.1029/2010GL044533.
Gómez-Pérez N and Wicht J (2010) Behavior of planetary dynamos under the influence of external magnetic fields: Application to Mercury and Ganymede. *Icarus* 209(1): 53–62.
Goossens S and Matsumoto K (2008) Lunar degree 2 potential Love number determination from satellite tracking data. *Geophysical Research Letters* 35: L02204. http://dx.doi.org/10.1029/2007GL031960.
Grasset O, Schneider J, and Sotin C (2008) A study of the accuracy of mass-radius relationships for silicate-rich and ice-rich planets up to 100 Earth masses. *Astrophysical Journal* 693: 722–733.
Greenwood JP, Itoh S, Sakamoto N, Warren PH, Taylor LA, and Yurimoto H (2011) Hydrogen isotope ratios in lunar rocks indicate delivery of cometary water to the Moon. *Nature Geoscience* 4: 79–81.
Grimm RE and Hess PC (1997) The crust of Venus. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1205–1244. Tucson, AZ: University of Arizona Press.
Grimm RE and Phillips RJ (1992) Anatomy of a Venusian hot-spot - Geology, gravity, and mantle dynamics of Eistla Regio. *Journal of Geophysical Research* 97(E10): 378–388.
Grott M, Baratoux D, Hauber E, et al. (2013) Long-term evolution of the Martian crust-mantle system. *Space Science Reviews* 174: 49–111.
Grott M and Wieczorek MA (2012) Density and lithospheric structure at Tyrrhena Patera, Mars, from gravity and topography data. *Icarus* 221: 43–52.
Gudkova TV, Lognonne P, Zharkov VN, and Raevsky SN (2014) On the scientific aims of the MISS seismic experiment. *Solar System Research* 48: 11–21.
Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of the Earth and Planetary Interiors* 142: 1–22.
Guest A and Smrekar SE (2007) New constraints on the thermal and volatile evolution of Mars. *Physics of the Earth and Planetary Interiors* 164: 161–176.
Halekas JS, Mitchell DL, Lin RP, et al. (2001) Mapping of crustal magnetic anomalies on the lunar near side by the Lunar Prospector electron reflectometer. *Journal of Geophysical Research* 106(E11): 27841–27852.
Harmon JK (2007) Radar imaging of Mercury. *Space Science Reviews* 132(2–4): 307–349.
Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's north pole. *Icarus* 149: 1–15.
Harmon JK, Slade MA, and Rice MS (2011) Radar imagery of Mercury's putative polar ice: 1999–2005 Arecibo results. *Icarus* 211: 37–50.
Hauck SA, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222: 713–728.
Hauck SA II, Margot J-L, Solomon SC, et al. (2013) The curious case of Mercury's internal structure. *Journal of Geophysical Research* 118: 1204–1220. http://dx.doi.org/10.1002/jgre.20091.
Hauck SA, Phillips RJ, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103(E6): 13,635–13,642.
Hauri EH, Weinreich T, Saal AE, Rutherford MC, and Van Orman JA (2011) High pre-eruptive water contents preserved in lunar melt inclusions. *Science* 333: 213–215.
Head JW III, Chapman CR, Strom RG, et al. (2011) Flood volcanism in the northern high latitudes of Mercury revealed by MESSENGER. *Science* 333: 1853–1856.
Head JW, Kreslavsky MA, and Pratt S (2002) Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *Journal of Geophysical Research* 107(E1). http://dx.doi.org/10.1029/2001JE001445.
Heiken G, Vaniman D, and French BM (1991) *The Lunar Sourcebook*. New York: Cambridge University Press.
Heimpel MH, Aurnou JM, Al-Shamali FM, and Gómex-Pérez N (2005) A numerical study of dynamo action as a function of spherical shell geometry. *Earth and Planetary Science Letters* 236: 542–557.
Hemley RJ (2006) Erskine Williamson, extreme conditions, and the birth of mineral physics. *Physics Today* 59: 50–56.
Hemley RJ and Ashcroft NW (1998) The revealing role of pressure in the condensed matter sciences. *Physics Today* 51: 26–32.
Herrick RR, Bills BG, and Hall SA (1989) Variations in effective compensation depth across Aphrodite Terra, Venus. *Geophysical Research Letters* 16: 543–546.
Herrick RR, Sharpton VL, Malin MC, Lyons SN, and Feely K (1997) Morphology and morphometry of impact craters. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1015–1046. Tucson, AZ: University of Arizona Press.

Heyner D, Schmitt D, Glassmeier KH, and Wicht J (2011a) Dynamo action in an ambient field. *Astronomische Nachrichten* 332(1): 36–42.
Heyner D, Wicht J, Gómex-Pérez N, Schmitt D, Auster HU, and Glassmeier KH (2011b) Evidence from numerical experiments for a feedback dynamo generating Mercury's magnetic field. *Science* 334: 1690–1693.
Hiesinger H, Head JW, Wolf U, Jaumann R, and Neukum G (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108(E7): 5065. http://dx.doi.org/10.1029/2002JE001985.
Hobbs BA (1987) Conductivity profiles from global data. *Pure and Applied Geophysics* 125: 393–407.
Hood LL (1986) Geophysical constraints on the lunar interior. In: Phillips RJ, Hartmann WK, and Taylor GJ (eds.) *Origin of the Moon*, pp. 397–409. Houston: Lunar and Planetary Institute.
Hood LL, Herbert F, and Sonett CP (1982) The deep lunar electrical conductivity profile – Structural and thermal inferences. *Journal of Geophysical Research* 87: 5311–5326.
Hood LL and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins: Two-dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.
Hood LL and Jones JH (1987) Geophysical constraints on lunar bulk composition and structure: A reassessment. In: *Journal of Geophysical Research*, 92, *Proceedings of the 17th Lunar and Planetary Science Conference, Part 2*, pp. E396–E410.
Hood LL, Mitchell DL, Lin RP, Acuña MH, and Binder AB (1999) Initial measurements of the lunar induced magnetic dipole moment using Lunar Prospector magnetometer data. *Geophysical Research Letters* 26: 2327–2330.
Hood LL and Sonett CP (1982) Limits on the lunar temperature profile. *Geophysical Research Letters* 9: 37–40.
Hood LL, Zakharian A, Halekas J, et al. (2001) Initial mapping and interpretation of lunar crustal magnetic anomalies using Lunar Prospector magnetometer data. *Journal of Geophysical Research* 106(E11): 27825–27839.
Hood LL and Zuber MT (2000) Recent refinements in geophysical constraints on lunar origin and evolution. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 397–409. Tucson, AZ: University of Arizona Press.
Humayun M, Nemchin A, Zanda B, et al. (2013) Origin and age of the earliest Martian crust from meteorite NWA 7533. *Nature* 29(3). http://dx.doi.org/10.1029/2001GL013853.
Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-neptunian objects. *Icarus* 185: 258–273.
Jagoutz E (1991) Chronology of SNC meteorites. *Space Science Reviews* 56: 13–22.
Jakosky BM and Phillips RJ (2001) Mars' volatile and climate history. *Nature* 412: 237–244.
Janle P and Meissner R (1986) Structure and evolution of the terrestrial planets. *Surveys in Geophysics* 8: 107–186.
Jaumann R, Hiesinger H, Anand M, et al. (2012) Geology, geochemistry, and geophysics of the Moon: Status of current understanding. *Planetary and Space Science* 74: 15–41.
Jehn R, Corral C, and Giampieri G (2004) Estimating Mercury's 88-day libration amplitude from orbit. *Planetary and Space Science* 52: 727–732.
Jolliff BL, Gaddis LR, Ryder G, et al. (2000a) New views of the Moon: Improved understanding through data integration. *EOS, Transactions American Geophysical Union* 81: 354–355, 349.
Jolliff BL, Gillis JJ, Haskin LA, Korotev RL, and Wieczorek MA (2000b) Major lunar crustal terranes: Surface expressions and crust-mantle origins. *Journal of Geophysical Research* 105: 4197–4216.
Katz RF, Spiegelman M, and Langmuir CH (2003) A new parameterization of hydrous mantle melting. *Geochemistry, Geophysics, Geosystems* 4: 1073. http://dx.doi.org/10.1029/2002GC000433.
Kaula WM (1979) The moment of inertia of Mars. *Geophysical Research Letters* 6: 194–196.
Kaula WM (1994) The tectonics of Venus. *Philosophical Transactions of the Royal Society of London A* 349: 345–355.
Kaula WM, Drake MJ, and Head JW (1986) The Moon. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 581–628. Tucson, AZ: University of Arizona Press.
Kaula WM, Lenardic A, Bindschadler DL, and Arkani-Hamed J (1997) Ishtar terra. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 879–900. Tucson, AZ: University of Arizona Press.
Kaula WM and Phillips RJ (1981) Quantitative tests for plate tectonics on Venus. *Geophysical Research Letters* 8: 1187–1190.
Kaula WM, Schubert G, Lingenfelter RE, Sjogren WL, and Wollenhaupt WR (1974) Apollo laser altimetry and inferences as to the lunar structure. In: *Proceedings of the 5th Lunar Planetary Science Conference*, vol. 3, pp. 3049–3058. Pergamon Press.
Kavner A, Duffy TS, and Shen G (2001) Phase stability and density of FeS at high pressures and temperatures: Implications for the interior structure of Mars. *Earth and Planetary Science Letters* 185: 25–33.
Keil K (2002) Geological history of asteroid 4 Vesta. In: Bottke WF, Paolicchi P, Binzel RP, and Cellino A (eds.) *Asteroids III*, pp. 573–584. Tucson, AZ: University of Arizona Press.
Khan A, Connolly JAD, Maclennon J, and Mosegaard K (2007) Joint inversion of seismic and gravity data for lunar composition and thermal state. *Geophysical Journal International* 168: 243–258.
Khan A, Maclennon J, Taylor SR, and Connolly JAD (2006) Are the Earth and the Moon compositionally alike? Inferences on lunar composition and implications for lunar origin and evolution from geophysical modeling. *Journal of Geophysical Research* 111(E05): E05005. http://dx.doi.org/10.1029/2005JE002608.
Khan A and Mosegaard K (2002) An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data. *Journal of Geophysical Research* 107(E6). http://dx.doi.org/10.1029/2001JE001658.
Khan A, Mosegaard K, and Rasmussen KL (2000) A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data. *Geophysical Research Letters* 27: 1591–1594.
Khan A, Mosegaard K, Williams JG, and Lognonne P (2004) Does the Moon possess a molten core? Probing the deep lunar interior using results from LLR and Lunar Prospector. *Journal of Geophysical Research* 109(E09): E09007. http://dx.doi.org/10.1029/2004JE002294.
Khurana KK, Jia X, Kivelson MG, Nimmo F, Schubert G, and Russell CT (2011) Evidence of a global magma ocean in Io's interior. *Science* 332: 1186–1189.
Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.
Kivelson MG, Bagenal F, Kurth WS, Neubauer FM, Paranicas C, and Saur J (2004) Magnetospheric interactions with satellites. In: Bagenal F, Dowling T, and Mckinnon W (eds.) *Jupiter, The Planet, Satellites, and Magnetosphere*, pp. 513–536. Cambridge, UK: Cambridge University Press.
Kivelson MG, Khurana KK, Russell CT, Volwerk M, Walker RJ, and Zimmer C (2000) Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.
Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104(A3): 4609–4626.
Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
Kleine T, Munker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.
Kleine T, Touboul M, Bourdon B, et al. (2009) Hf-W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochimica et Cosmochimica Acta* 73: 5150–5188.
Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars' seismicity. *Journal of Geophysical Research* 111: E11006. http://dx.doi.org/10.1029/2006JE002708.
Konopliv AS, Asmar SW, Carranza E, Sjogren WL, and Yuan DN (2001) Recent gravity models as a result of the Lunar Prospector mission. *Icarus* 150: 1–18.
Konopliv AS, Asmar SW, Folkner WM, et al. (2011) Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters. *Icarus* 211: 401–428.
Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.
Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from Lunar Prospector. *Science* 281: 1476–1480.
Konopliv AS, Park RS, Yuan D-N, et al. (2013) The JPL lunar gravity field to spherical harmonic degree 660 from the GRAIL Primary Mission. *Journal of Geophysical Research* 118: 1415–1434. http://dx.doi.org/10.1002/jgre.20097.
Konopliv AS and Yoder CF (1996) Venusian k_2 tidal Love number from Magellan and PVO tracking data. *Geophysical Research Letters* 23: 1857–1860.
Konopliv AS, Yoder CF, Standish EM, Yuan DN, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 23–50.
Korotev RL (2005) Lunar geochemistry as told by lunar meteorites. *Chemie der Erde – Geochemistry* 65: 297–346.
Kovach RL and Chyba CF (2001) Seismic detectability of a subsurface ocean on Europa. *Icarus* 150: 279–287.
Kuchynka P, Folkner WM, Konopliv AS, et al. (2014) New constraints on Mars rotation determined from radiometric tracking of the Opportunity Mars exploration Rover. *Icarus* 229: 340–347.

Kucinskas AB and Turcotte DL (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112: 104–116.

Kuskov OL (1995) Constitution of the Moon: 3. Composition of the middle mantle from seismic data. *Physics of the Earth and Planetary Interiors* 90: 55–74.

Kuskov OL (1997) Constitution of the Moon: 4. Composition of the mantle from seismic data. *Physics of the Earth and Planetary Interiors* 102: 239–257.

Kuskov OL and Fabrichnaya OB (1995) Constitution of the Moon: 2. Composition and seismic properties of the lower mantle. *Physics of the Earth and Planetary Interiors* 83: 197–216.

Kuskov OL and Kronrod VA (1998) Constitution of the Moon: 5. Constraints on composition, density, temperature, and radius of a core. *Physics of the Earth and Planetary Interiors* 107: 285–306.

Kuvshinov A and Olsen N (2006) A global model of mantle conductivity derived from 5 years of CHAMP, Ørsted, and SAC-C magnetic data. *Geophysical Research Letters* 33, L18301.

Lahiri BN and Price AT (1939) Electromagnetic induction in non-uniform conductors, and the determination of the conductivity of the earth from terrestrial magnetic variations. *Philosophical Transactions of the Royal Society of London A* 237: 509–540.

Lainey V, Dehant V, and Pätzold M (2007) First numerical ephemerides of the Martian moons. *Astronomy and Astrophysics* 465: 1075–1084.

Langevin Y (1997) The regolith of Mercury: Present knowledge and implications for the Mercury Orbiter mission. *Planetary and Space Science* 45: 31–37.

Lawrence DJ, Feldman WC, Goldsten JO, et al. (2013) Evidence for water ice near Mercury's north pole from MESSENGER Neutron Spectrometer measurements. *Science* 339: 292–296.

Lay T, Heinz D, Ishil M, et al. (2005) Multidisciplinary impact of the deep mantle phase transition in perovskite structure. *Eos, Transactions of the American Geophysical Union* 86: 1,5.

Lee S, Zanolin M, Thode AM, Pappalardo RT, and Makris NC (2003) Probing Europa's interior with natural sound sources. *Icarus* 165: 144–167.

Lemoine FG, Goossens S, Sabaka TJ, et al. (2013) High-degree gravity models from GRAIL primary mission data. *Journal of Geophysical Research* 118: 1676–1698. http://dx.doi.org/10.1002/jgre.20118.

Lemoine FG, Smith DE, Rowlands DD, et al. (2001) An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor. *Journal of Geophysical Research* 106: 23,359–23,376.

Liebermann RC (2005) The future of high-pressure mineral physics. *Eos, Transactions of the American Geophysical Union* 86(40): 365–373.

Lillis RJ, Robbins S, Manga M, Halekas JS, and Frey HV (2013) Time history of the Martian dynamo from crater magnetic field analysis. *Journal of Geophysical Research* 118: 1488–1511. http://dx.doi.org/10.1002/jgre.20105.

Lin RP, Mitchell DL, Curtis DW, et al. (1998) Lunar surface magnetic fields and their interaction with the solar wind: Results from Lunar Prospector. *Science* 281: 1480–1484.

Lingenfelter RE and Schubert G (1973) Evidence for convection in planetary interiors from first-order topography. *Earth, Moon, and Planets* 7: 172–180.

Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 571–604.

Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.

Lognonné P, Giardini D, Banerdt B, et al. (2000) The NetLander very broad band seismometer. *Planetary and Space Science* 48: 1289–1302.

Lognonné P and Mosser B (1993) Planetary seismology. *Surveys in Geophysics* 14: 239–302.

Longhi J, Knittle E, Holloway JR, and Wänke H (1992) The bulk composition, mineralogy and internal structure of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 184–208. Tucson, AZ: University of Arizona Press.

Malavergne V, Toplis MJ, Berthet S, and Jones J (2010) Highly reducing conditions during core formation on Mercury: Implications for internal structure and the origin of a magnetic field. *Icarus* 206: 199–209.

Manglik A, Wicht J, and Christensen UR (2010) A dynamo model with double diffusive convection for Mercury's core. *Earth and Planetary Science Letters* 289: 619–628.

Margot JL, Campbell DB, Jurgens RF, and Slade MA (1999) Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* 284: 1658–1660.

Margot JL, Peale SJ, Jurgens RF, Slade MA, and Holin IV (2007) Large longitude libration of Mercury reveals a molten core. *Science* 316: 710–714.

Margot JL, Peale SJ, Solomon SC, et al. (2012) Mercury's moment of inertia from spin and gravity data. *Journal of Geophysical Research* 117: E00L09. http://dx.doi.org/10.1029/2012JE004161.

Marty JC, Duron J, Balmino G, Dehant V, Rosenblatt P, and Hoolst TV (2009) Martian gravity field model and its time variations. *Planetary and Space Science* 57: 350–363.

Marty B and Marti K (2002) Signatures of early differentiation of Mars. *Earth and Planetary Science Letters* 196: 251–263.

Matsumoto K, Goossens S, Ishihara Y, et al. (2010) An improved lunar gravity field model from SELENE and historical tracking data: Revealing the farside gravity features. *Journal of Geophysical Research* 115: E06007. http://dx.doi.org/10.1029/2009JE003499.

Mccubbin FM, Hauri EH, Elardo SM, Vander Kaaden KE, Wang J, and Shearer CK Jr (2012a) Hydrous melting of the martian mantle produced both depleted and enriched shergottites. *Geology* 40: 683–686.

Mccubbin FM, Riner MA, Vander Kaaden KE, and Burkemper LK (2012b) Is Mercury a volatile-rich planet? *Geophysical Research Letters* 39. http://dx.doi.org/10.1029/2012GL051711.

Mccubbin FM, Smirnov A, Nekvasil H, Wang J, Hauri EH, and Lindsley DH (2010a) Hydrous magmatism on Mars: A source of water for the surface and subsurface during the Amazonian. *Earth and Planetary Science Letters* 292: 132–138.

Mccubbin FM, Steele A, Hauri EH, Nekvasil H, Yamashita S, and Hemley RJ (2010b) Nominally hydrous magmatism on the Moon. *Proceedings of the National Academy of Sciences of the United States of America* 107: 11223–11228.

McDonough WF and Sun S-S (1995) Composition of the Earth. *Chemical Geology* 120: 223–253.

McKenzie D, Ford PG, Johnson C, et al. (1992) Features on Venus generated by plate boundary processes. *Journal of Geophysical Research* 97(E8): 13533–13544.

McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130: 198–216.

Mckinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology Geophysics, Atmosphere, and Solar Wind Environment*, pp. 969–1014. Tucson, AZ: University of Arizona Press.

Mcsween HY (1994) What we have learned about Mars from SNC meteorites. *Meteoritics* 29: 757–779.

Mcsween HY, Grove TL, Lentz RCF, et al. (2001) Geochemical evidence for magmatic water within Mars from pyroxenes in the Shergotty meteorite. *Nature* 409: 487–490.

Milbury C, Schubert G, Raymond CA, Smrekar SE, and Langlais B (2012) The history of Mars' dynamo as revealed by modeling magnetic anomalies near Tyrrhenus Mons and Syrtis Major. *Journal of Geophysical Research* 117: E10007. http://dx.doi.org/10.1029/2012JE004099.

Mitrofanov IG, Sanin AB, Boynton WV, et al. (2010) Hydrogen mapping of the Lunar south pole using the LRO Neutron Detector Experiment LEND. *Science* 330: 483–486.

Mocquet A, Rosenblatt P, Dehant V, and Verhoeven O (2011) The deep interior of Venus, Mars, and the Earth: A brief review and the need for planetary surface-based measurements. *Planetary and Space Science* 59: 1048–1061.

Mocquet A, Vacher P, Grasset O, and Sotin C (1996) Theoretical seismic models of Mars: The importance of the iron content of the mantle. *Planetary and Space Science* 44: 1251–1268.

Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regressions of highland geoid and topography. *Icarus* 128: 415–428.

Moore WB, Schubert G, Anderson JD, and Spencer JR (2006) The interior of Io. In: Lopes RMC and Spencer JR (eds.) *Io after Galileo*, pp. 89–108. Chichester, UK: Springer-Verlag.

Morard G and Katsura T (2010) Pressure-temperature cartography of Fe-S-Si immiscible system. *Geochimica et Cosmochimica Acta* 74: 3659–3667.

Muhleman DO, Grossman AW, and Butler BJ (1995) Radar investigation of Mars, Mercury and Titan. *Annual Review of Earth and Planetary Sciences* 23: 337–374.

Murakami M, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in $MgSiO_3$. *Science* 304: 855–858.

Nakamura Y (1983) Seismic velocity structure of the lunar mantle. *Journal of Geophysical Research* 88: 677–686.

Nakamura Y (2005) Farside deep moonquakes and deep interior of the moon. *Journal of Geophysical Research* 110: E01001. http://dx.doi.org/10.1029/2004JE002332.

Nakamura Y, Latham GV, and Dorman HJ (1982) Apollo lunar seismic experiment final summary. *Journal of Geophysical Research* 87: A117–A123.

Nakamura Y, Latham G, Lammlein D, Ewing M, Duennebier F, and Dorman J (1974) Deep lunar interior inferred from recent seismic data. *Geophysical Research Letters* 1: 137–140.

Namiki N, Iwata T, Matsumoto K, et al. (2009) Farside gravity field of the Moon from four-way Doppler measurements of SELENE (Kaguya). *Science* 323: 900–905.

Ness NF (1979) The magnetic fields of Mercury, Mars, and Moon. *Annual Review of Earth and Planetary Sciences* 7: 249–288.

Neumann GA, Cavanaugh JF, Sun X, et al. (2013) Bright and dark polar deposits on Mercury: Evidence for surface volatiles. *Science* 339: 296–300.
Neumann GA, Zuber MT, Smith DE, and Lemoine FG (1996) The lunar crust: Global structure and signature of major basins. *Journal of Geophysical Research* 101: 16,841–16,863.
Neumann GA, Zuber MT, Wieczorek MA, Mcgovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E08002. http://dx.doi.org/10.1029/2004JE002262.
Nimmo F and McKenzie D (1998) Volcanism and tectonics on Venus. *Annual Review of Earth and Planetary Sciences* 26: 23–51.
Nimmo F and Stevenson DJ (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11979.
Nimmo F and Stevenson DJ (2001) Estimates of Martian crustal thickness from viscous relaxation of topography. *Journal of Geophysical Research* 106: 5085–5098.
Nimmo F and Tanaka K (2005) Early crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161.
Nishida K, Terasaki H, Ohtani E, and Suzuki A (2008) The effect of sulfur content on density of the liquid Fe-S at high pressure. *Physics and Chemistry of Minerals* 35: 417–423.
Nittler LR, Starr RD, Weider SZ, et al. (2011) The major-element composition of Mercury's surface from MESSENGER X-ray spectrometry. *Science* 333: 1847–1850.
Nozette S, Rustan P, Pleasance LP, et al. (1994) The Clementine mission to the Moon: Scientific overview. *Science* 266: 1835–1839.
Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ perovskite in Earth's D" layer. *Nature* 430: 445–448.
Olsen N (1999) Induction studies with satellite data. *Surveys in Geophysics* 20: 309–340.
Paige DA, Siegler MA, Harmon JK, et al. (2013) Thermal stability of volatiles in the north polar region of Mercury. *Science* 339: 300–303.
Panning M, Lekic V, Manga M, Cammarano F, and Romanowicz B (2006) Long-period seismology on Europa: 2. Predicted seismic response. *Journal of Geophysical Research* 111: E12008. http://dx.doi.org/10.1029/2006JE002712.
Pauer M and Breuer D (2008) Constraints on the maximum crustal density from gravity-topography modeling: Applications to the southern highlands of Mars. *Earth and Planetary Science Letters* 276: 253–261.
Peale SJ (1976) Inferences from the dynamical history of Mercury's rotation. *Icarus* 28: 459–467.
Peale SJ (1988) The rotational dynamics of Mercury and the state of its core. In: Vilas F, Chapman R, and Matthews MS (eds.) *Mercury*, pp. 461–493. Tucson, AZ: University of Arizona Press.
Peplowski PN, Evans LG, Hauck SA II, et al. (2011) Radioactive elements on Mercury's surface from MESSENGER: Implications for the planet's formation and evolution. *Science* 333: 1850–1852.
Phillips RJ (1994) Estimating lithospheric properties at Atla Regio, Venus. *Icarus* 112: 147–170.
Phillips JL and Russell CT (1987) Upper limit on the intrinsic magnetic field of Venus. *Journal of Geophysical Research* 92(A3): 2253–2263.
Pieters CM and Tompkins S (1999) Tsiolkovsky crater: A window into crustal processes on the lunar farside. *Journal of Geophysical Research* 104: 21,935–21,950.
Platzman GW (1984) Planetary energy balance for tidal dissipation. *Reviews of Geophysics and Space Physics* 22: 73–84.
Poirier JP (1994) Light elements in the Earth's outer core: A critical review. *Earth and Planetary Science Letters* 85: 319–337.
Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active South Pole of Enceladus. *Science* 311: 1393–1401.
Purucker ME, Johnson CL, Winslow RM, et al. (2012) Evidence for a crustal magnetic signature on Mercury from MESSENGER magnetometer observations. In: *Lunar Planetary Science Conference 43,* abstract #1297.
Quesnel Y, Langlais B, and Sotin C (2007) Local inversion of magnetic anomalies: Implications for Mars' crustal evolution. *Planetary and Space Science* 55: 258–269.
Quesnel Y, Sotin C, Langlais B, et al. (2009) Serpentinization of the martian crust during Noachian. *Earth and Planetary Science Letters* 277: 184–193.
Rappaport NJ, Konopliv AS, Kucinskas AB, and Ford PG (1999) An improved 360 degree and order model of Venus topography. *Icarus* 139: 19–31.
Rauer H, Catala C, Aerts C, et al. (2014) The PLATO 2.0 mission. *Experimental Astronomy.* http://dx.doi.org/10.1007/s10686-014-9383-4.
Reasenberg RD (1977) The moment of inertia and isostasy of Mars. *Journal of Geophysical Research* 82: 369–375.
Righter K and Drake MJ (1996) Core formation in Earth's Moon, Mars, and Vesta. *Icarus* 124: 513–529.
Righter K and Drake MJ (1997) A magma ocean on Vesta: Core formation and petrogenesis of eucrites and diogenites. *Meteoritics and Planetary Science* 32: 929–944.
Rikitake T (1966) *Electromagnetism and the Earth's Interior*. Amsterdam: Elsevier.
Ringwood AE (1970) The system Mg_2SiO_4-Fe_2SiO_4 at high pressures and temperatures. *Physics of the Earth and Planetary Interiors* 3: 89–108.
Ringwood AE (1975) *Composition and Petrology of the Earth's Mantle*. New York: McGraw-Hill.
Ringwood AE (1977) Composition of the core and implications for the origin of the Earth. *Geochemical Journal* 11: 111–135.
Ringwood AE (1979) *On the Origin of the Earth and Moon*. New York: Springer-Verlag.
Rivoldini A and Van Hoolst T (2013) The interior structure of Mercury constrained by the low-degree gravity field and the rotation of Mercury. *Earth and Planetary Science Letters* 377–378: 62–72.
Rivoldini A, Van Hoolst T, and Verhoeven O (2009) The interior structure of Mercury and its core sulfur content. *Icarus* 201: 12–30.
Rivoldini A, Van Hoolst T, Verhoeven O, Mocquet A, and Dehant V (2011) Geodesy constraints on the interior structure and composition of Mars. *Icarus* 213: 451–472.
Roberts RG (1986) Global electromagnetic induction. *Surveys in Geophysics* 8: 339–374.
Roberts JH and Zhong S (2006) Degree-1 convection in the Martian mantle and the origin of the hemispheric dichotomy. *Journal of Geophysical Research* 111: E06013. http://dx.doi.org/10.1029/2005JE002668.
Russell CT (1980) Planetary magnetism. *Reviews of Geophysics* 18: 77–106.
Russell CT, Coleman PJ Jr., Fleming BK, et al. (1975) The fine-scale lunar magnetic field. In: *Proceedings of the 6th Lunar Science Conference,* vol. 3, pp. 2955–2969. Cambridge, MA: The MIT Press.
Russell CT, Coleman PJ Jr., Lichtenstein BR, Schubert G, and Sharp LR (1973) Subsatellite measurements of the lunar magnetic field. In: *Proceedings of the 4th Lunar Science Conference,* vol. 3, pp. 2833–2845. Cambridge, MA: The MIT Press.
Russell CT, Raymond CA, Coradini A, et al. (2012) Dawn at Vesta: Testing the protoplanetary paradigm. *Science* 336: 684–686.
Saal AE, Hauri EH, Cascio ML, Van Orman JA, Rutherford MC, and Cooper RF (2008) Volatile content of lunar volcanic glasses and the presence of water in the Moon's interior. *Nature* 454: 192–195.
Sandwell DT and Schubert G (1992a) Evidence for retrograde lithospheric subduction on Venus. *Science* 257: 766–770.
Sandwell DT and Schubert G (1992b) Flexural ridges, trenches, and outer rises around coronae on Venus. *Journal of Geophysical Research* 97(E10): 16,069–16,083.
Sanloup C and Fei Y (2004) Closure of the Fe-S-Si liquid miscibility gap at high pressure. *Physics of the Earth and Planetary Interiors* 147: 57–65.
Sanloup C, Guyot F, Gillet P, Fiquet G, Mezouar M, and Martinez I (2000) Density measurements of liquid Fe-S alloys at high-pressure. *Geophysical Research Letters* 27: 811–814.
Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of the Earth and Planetary Interiors* 112: 43–54.
Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus – What are they telling us. *Journal of Geophysical Research* 97(E8): 13,257–13,301.
Schenk P, O'Brien DP, Marchi S, et al. (2012) The geologically recent giant impact basins at Vesta's south. *Science* 336: 694–697.
Scholten F, Oberst J, Matz K-D, et al. (2012) GLD100: The near-global lunar 100 m raster DTM from LROC WAC stereo image data. *Journal of Geophysical Research* 117: E00H17. http://dx.doi.org/10.1029/2011JE003926.
Schubert G, Anderson JD, Spohn T, and Mckinnon WB (2004) Interior composition, structure and dynamics of the Galilean satellites. In: Bagenal F, Dowling TE, and Mckinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 281–306. Cambridge: Cambridge University Press.
Schubert G, Anderson JD, Travis BJ, and Palguta J (2007) Enceladus: Present internal structure and differentiation by early and long term radiogenic heating. *Icarus* 188: 345–355.
Schubert G and Lingenfelter RE (1973) Martian centre of mass – Centre of figure offset. *Nature* 242: 251–252.
Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*', pp. 429–460. Tucson, AZ: University of Arizona Press.
Schubert G, Russell CT, and Moore WB (2000) Geophysics-timing of the Martian dynamo. *Nature* 408: 666–667.
Schubert G and Schwartz K (1969) A theory for the interpretation of lunar surface magnetometer data. *The Moon* 1: 106–117.

Schubert G, Solomatov VS, Tackley PJ, and Turcotte DL (1997) Mantle convection and the thermal evolution of Venus. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1245–1287. Tucson, AZ: University of Arizona Press.

Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.

Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.

Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.

Schubert G, Zhang K, Kivelson MG, and Anderson JD (1996) The magnetic field and internal structure of Ganymede. *Nature* 384: 544–545.

Schuster A (1889) The diurnal variation of terrestrial magnetism. *Philosophical Transactions of the Royal Society of London A* 180: 467–518.

Seager S, Kuchner M, Hier-Majumder CA, and Militzer B (2007) Mass-radius relationships for solid exoplanets. *Astrophysical Journal* 669: 1279–1297.

Seidelmann PK, Abalkin VK, Bursa M, et al. (2002) Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celestial Mechanics and Dynamical Astronomy* 82: 83–110.

Seidelmann PK, Archinal BA, A'Hearn MF, et al. (2005) Report of the IAU/IAG working group on cartographic coordinates and rotational elements: 2003. *Celestial Mechanics and Dynamical Astronomy* 91: 203–215.

Seidelmann PK, Archinal BA, A'Hearn MF, et al. (2007) Report of the IAU/IAG working group on cartographic coordinates and rotational elements: 2006. *Celestial Mechanics and Dynamical Astronomy* 98: 155–180.

Seiff A (1983) Thermal structure of the atmosphere of Venus. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 154–158. Tucson, AZ: University of Arizona Press.

Senske DA, Schaber GG, and Stofan ER (1992) Regional topographic rises on Venus – Geology of Western Eistla Regio and comparison to Beta Regio and Atla Regio. *Journal of Geophysical Research* 97(E8): 13395–13420.

Sharp ZD, Shearer CK, Mckeegan KD, Barnes JD, and Wang YQ (2010) The chlorine isotope composition of the Moon and implications for an anhydrous mantle. *Science* 329: 1050–1053.

Shea EK, Weiss BP, Cassata WS, et al. (2012) A long-lived lunar core dynamo. *Science* 335: 453–456.

Shearer CK, Hess PC, Wieczorek MA, et al. (2006) Thermal and magmatic evolution of the Moon. *Reviews in Mineralogy and Geochemistry* 60: 365–518.

Shimizu H, Matsushima M, Takahashi F, Shibuya H, and Tsunakawa H (2013) Constraint on the lunar core size from electromagnetic sounding based on magnetic field observations by orbiting satellite. *Icarus* 222: 32–43.

Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid/topography admittance of Venus. *Science* 264: 798–803.

Sjogren WL, Banerdt WB, Chodas PW, et al. (1997) The Venus gravity field and other geodetic parameters. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1125–1162. Tucson, AZ: University of Arizona Press.

Slade MA, Butler BJ, and Muhleman DO (1992) Mercury's radar imaging: Evidence for polar ice. *Science* 258: 635–640.

Sloan ED (2003) Fundamental principles and applications of natural gas hydrates. *Nature* 426: 353–359.

Smith JV, Anderson AT, Newton RC, et al. (1970) Petrologic history of the Moon inferred from petrography, mineralogy, and petrogenesis of Apollo 11 rocks. In: *Proceedings of the Apollo 11 Lunar Science Conference,* vol. 1, pp. 897–925. New York: Pergammon Press.

Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.

Smith DE, Sjogren WL, Tyler GL, Balmino G, Lemoine FG, and Konopliv AS (1999a) The gravity field of Mars: Results from Mars Global Surveyor. *Science* 286: 94–97.

Smith DE and Zuber MT (1996) The shape of Mars and the topographic signature of the hemispheric dichotomy. *Science* 271: 184–187.

Smith DE, Zuber MT, Phillips RJ, et al. (2010) The equatorial shape and gravity field of Mercury from MESSENGER flybys 1 and 2. *Icarus* 209: 88–100.

Smith DE, Zuber MT, Phillips RJ, et al. (2012) Gravity field and internal structure of Mercury from MESSENGER. *Science* 336: 214–217.

Smith DE, Zuber MT, Solomon SC, et al. (1999b) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.

Smrekar SE (1994) Evidence for active hotspots on Venus from analysis of Magellan gravity data. *Icarus* 112: 2–26.

Smrekar SE, Stofan ER, Mueller N, et al. (2010) Recent hot-spot volcanism on Venus from VIRTIS emissivity data. *Science* 328: 605–608.

Snyder Hale A and Hapke B (2002) A time-dependent model of radiative and conductive thermal energy transport in planetary regoliths with applications to the Moon and Mercury. *Icarus* 156: 318–334.

Sobolev SV and Babeyko AY (1994) Modeling of mineralogical composition, density, and elastic wave velocity in anhydrous magmatic rocks. *Surveys in Geophysics* 15: 515–544.

Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008. http://dx.doi.org/10.1029/2005JE002520.

Sohl F and Spohn T (1997) The interior structure of Mars: Implications from SNC meteorites. *Journal of Geophysical Research* 102: 1613–1635.

Sohl F, Wagner FW, Hussmann H, and Grott M (2009) Terrestrial planets and satellites: Planetary interiors. In: Martienssen W and Triimper J (eds.) *Landolt–Bornstein – Numerical Data and Functional Relationships in Science and Technology. Astronomy Astrophysics, and Cosmology – Subvolume B*, vol. VI/4B, pp. 200–224. Berlin, Heidelberg, New York: Springer-Verlag.

Solomon SC (2003) Mercury: The enigmatic innermost planet. *Earth and Planetary Science Letters* 216: 441–455.

Solomon SC, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.

Solomon SC, Mcnutt RL Jr., Watters TR, et al. (2008) Return to Mercury: A global perspective on MESSENGER's first Mercury flyby. *Science* 321: 59–62.

Sonett CP (1982) Electromagnetic induction in the Moon. *Reviews of Geophysics and Space Physics* 20: 411–455.

Sonett CP, Colburn DS, Dyal P, et al. (1971) Lunar electrical conductivity profile. *Nature* 230: 359–362.

Sonett CP, Smith BF Colburn DS Schubert G and Schwartz K (1972) The induced magnetic field of the Moon: Conductivity profiles and inferred temperatures. In *Proceedings of the 3rd Lunar Science Conference, Geochim. Cosmochim. Acta, Suppl. 3*, pp. 2309–2336. The MIT Press.

Sotin C, Grasset O, and Mocquet A (2007) Mass-radius curve for extrasolar earth-like planets and ocean planets. *Icarus* 191: 337–351.

Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. *Science* 311: 1401–1405.

Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–236.

Spohn T, Acuña MH, Breuer D, et al. (2001a) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.

Spohn T, Sohl F, and Breuer D (1998) Mars. *Astronomy and Astrophysics Review* 8: 181–235.

Spohn T, Sohl F, Wieczerkowski K, and Conzelmann V (2001b) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary and Space Science* 49: 1561–1570.

Sprague AL, Kozlowski RWH, Witteborn FC, Cruikshank DP, and Wooden DH (1994) Mercury: Evidence for anorthosite and basalt from mid-infrared (7.3–13.5 um) spectroscopy. *Icarus* 109: 156–167.

Stacey FD (1977) Applications of thermodynamics to fundamental Earth physics. *Geophysical Surveys* 3: 175–204.

Stacey FD, Brennan BJ, and Irvine RD (1981) Finite strain theories and comparisons with seismological data. *Geophysical Surveys* 4: 189–232.

Stamenković V, Noack L, Breuer D, and Spohn T (2012) The influence of pressure-dependent viscosity on the thermal evolution of super-Earths. *Astrophysical Journal* 748(41): 1–22.

Stanley S, Bloxham J, Hutchison WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.

Stanley S, Elkins-Tanton L, Zuber MT, and Parmentier EM (2008) Mars' paleomagnetic field as the result of a single-hemisphere dynamo. *Science* 321: 1822–1825.

Stanley S and Mohammadi A (2008) Effects of an outer thin stably stratified layer on planetary dynamos. *Physics of the Earth and Planetary Interiors* 168: 179–190.

Steinke T (2012). *Modeling Mercury's Interior Structure and Tidal Deformation*. Bachelor's Thesis (in German), Karlsruhe Institute of Technology, Karlsruhe.

Stevenson DJ (1982) Volcanism and igneous processes in icy satellites. *Nature* 298: 142–144.

Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.

Stofan ER, Hamilton VE, Janes DM, and Smrekar SE (1997) Coronae on Venus: Morphology and origin. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 931–965. Tucson, AZ: University of Arizona Press.

Strom RG (1987) *Mercury: The Elusive Planet*. Washington, DC; London: Smithsonian Institution Press.

Strom RG (1997) Mercury: An overview. *Advances in Space Research* 19: 1471–1485.
Strom RG, Banks ME, Chapman CR, et al. (2011) Mercury crater statistics from MESSENGER flybys: Implications for stratigraphy and resurfacing history. *Planetary and Space Science* 59: 1960–1967.
Surkov YA (1983) Studies of Venus rocks by Veneras 8, 9, and 10. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 154–158. Tucson, AZ: University of Arizona Press.
Swift DC, Eggert JH, Hicks DG, et al. (2012) Mass-radius relationships for exoplanets. *Astrophysical Journal* 744: 59-1–59-10.
Takahashi F and Matsushima M (2006) Dipolar and non-dipolar dynamos in a thin shell geometry with implications for the magnetic field of Mercury. *Geophysical Research Letters* 33: L10202. http://dx.doi.org/10.1029/2006GL025792.
Tarits P (1994) Electromagnetic studies of global geodynamic processes. *Surveys in Geophysics* 15: 209–238.
Toksöz MN, Dainty AM, Solomon SC, and Anderson KR (1974) Structure of the Moon. *Reviews of Geophysics* 12: 539–567.
Tsuchiya T, Tsuchiya J, Umemoto K, and Wentzcovitch RM (2004) Phase transition in $MgSiO_3$ perovskite in the earth's lower mantle. *Earth and Planetary Science Letters* 224: 241–248.
Tsuno K, Frost DJ, and Rubie DC (2011) The effects of nickel and sulphur on the core-mantle partitioning of oxygen in Earth and Mars. *Physics of the Earth and Planetary Interiors* 185: 1–12.
Turcotte DL (1993) An episodic hypothesis for Venusian tectonics. *Journal of Geophysical Research* 98(E9): 17,061–17,068.
Turcotte DL and Schubert G (2002) *Geodynamics*. Cambridge: Cambridge University Press.
Valencia D and O'Connell RJ (2009) Convection scaling and subduction on Earth and super-Earths. *Earth and Planetary Science Letters* 286: 492–502.
Valencia D, O'Connell RJ, and Sasselov DD (2006) Internal structure of massive terrestrial planets. *Icarus* 181: 545–554.
Valencia D, Sasselov DD, and O'Connell RJ (2007) Radius and structure models of the first super-earth planet. *Astrophysical Journal* 656: 545–551.
Van Hemelrijck E and Vercheval J (1981) Some aspects of the solar radiation incident at the top of the atmospheres of Mercury and Venus. *Icarus* 48: 167–179.
Van Hoolst T, Dehant V, Roosbeek F, and Lognonné P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 161: 281–296.
Van Hoolst T and Jacobs C (2003) Mercury's tides and interior structure. *Journal of Geophysical Research* 108(E11): 5121. http://dx.doi.org/10.1029/2003JE002126.
Van Thienen P, Rivoldini A, Van Hoolst T, and Lognonné P (2006) A top-down origin for martian mantle plumes. *Icarus* 185: 197–210.
Vasavada AR, Paige DA, and Wood SE (1999) Near-surface temperatures on Mercury and the Moon and the stability of polar ice deposits. *Icarus* 141: 179–193.
Verhoeven O, Rivoldini A, Vacher P, et al. (2005) Interior structure of terrestrial planets: Modelling Mars' mantle and its electromagnetic, geodetic and seismic properties. *Journal of Geophysical Research* 110: E04009. http://dx.doi.org/10.1029/2004JE002271.
Vilas F (1985) Mercury: Absence of crystalline Fe^{2+} in the regolith. *Icarus* 64: 133–138.
Vilim R, Stanley S, and Hauck SA (2010) Iron snow zones as a mechanism for generating Mercury's weak observed magnetic field. *Journal of Geophysical Research* 115(E11). http://dx.doi.org/10.1029/2009JE003528.
Voorhies CV, Sabaka TJ, and Purucker M (2002) On magnetic spectra of Earth and Mars. *Journal of Geophysical Research* 107(E6). http://dx.doi.org/10.1029/2001JE001534.
Wagner FW (2014) *The Physical State of Rocky Exoplanet Interiors*. Dissertation, Westphalian Wilhelms-University, Munster.
Wagner FW, Sohl F, Hussmann H, Grott M, and Rauer H (2011) Interior structure models of solid exoplanets using material laws in the infinite pressure limit. *Icarus* 214: 366–377.
Wagner FW, Tosi N, Sohl F, Rauer H, and Spohn T (2012) Rocky super-Earth interiors: Structure and internal dynamics of CoRoT-7b and Kepler-10b. *Astronomy and Astrophysics* 541: 1–11.
Wänke H (1981) Constitution of terrestrial planets. *Philosophical Transactions of the Royal Society of London A* 303: 287–302.
Wänke H (1991) Chemistry, accretion, and evolution of Mars. *Space Science Reviews* 56: 1–8.
Wänke H and Dreibus G (1994) Chemistry and accretion history of Mars. *Philosophical Transactions of the Royal Society of London A* 349: 285-293.
Wänke H and Dreibus G (1988) Chemical composition and accretion history of terrestrial planets. *Philosophical Transactions of the Royal Society of London A* 325: 545–557.
Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240.
Wasson JT (1988) The building stones of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.
Watters TR and McGovern PJ (2005) Hemispheres apart: The martian crustal dichotomy. *Eos, Transactions of the American Geophysical Union* 86(5): 46–47.
Weber RC, Lin P-Y, Garnero EJ, Williams Q, and Lognonne P (2011) Seismic detection of the lunar core. *Science* 331: 309–312.
Weitz CM and Basilevsky AT (1993) Magellan observations of the Venera and Vega landing site regions. *Journal of Geophysical Research* 98(E9): 17,069–17,097.
Wessel P and Smith WHF (1991) Free software helps map and display data. *Eos, Transactions of the American Geophysical Union* 72: 441,445–441,446.
Wicht J, Mandea M, Takahashi F, Christensen UR, Matsushima M, and Langlais B (2007) The origin of Mercury's internal magnetic field. *Space Science Reviews* 132(2–4): 261–290.
Wieczorek MA, Jolliff BL, Khan A, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364.
Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339: 671–675.
Wieczorek MA and Phillips RJ (1998) Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *Journal of Geophysical Research* 103: 1715–1724.
Wieczorek MA and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.
Wieczorek MA and Phillips RJ (2000) The Procellarum KREEP Terrane: Implications for mare volcanism and lunar evolution. *Journal of Geophysical Research* 105: 20,417–20,430.
Wieczorek MA and Zuber MT (2001) The composition and origin of the lunar crust: Constraints from central peaks and crustal thickness modeling. *Geophysical Research Letters* 28: 4023–4026.
Wieczorek MA and Zuber MT (2004) Thickness of the martian crust: Improved constraints from geoid-to-topography ratios. *Journal of Geophysical Research* 109: E01009. http://dx.doi.org/10.1029/2003JE002153.
Wieczorek MA, Zuber MT, and Phillips RJ (2001) The role of magma buoyancy on the eruption of lunar basalts. *Earth and Planetary Science Letters* 185: 71–83.
Wilhelms DE (1976) Mercurian volcanism questioned. *Icarus* 28: 551–558.
Wilhelms DE and Squyres SW (1984) The Martian hemispheric dichotomy may be due to a giant impact. *Nature* 309: 138–140.
Williams JG (2007) A scheme for lunar inner core detection. *Geophysical Research Letters* 34: L03202. http://dx.doi.org/10.1029/2006GL028185.
Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27,933–27,968.
Williamson ED and Adams LH (1923) Density distribution in the Earth. *Journal of the Washington Academy of Sciences* 13: 413–423.
Wood BJ (1993) Carbon in the core. *Earth and Planetary Science Letters* 117: 593–607.
Wood JA, Anderson DL, Buck WR, et al. (1981) Geophysical and cosmochemical constraints on properties of mantles of the terrestrial planets. In: Project BVS (ed.) *Basaltic Volcanism on the Terrestrial Planets*, pp. 633–699. New York: Pergamon Press, Inc.
Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites and a geophysical model of the Moon. In: *Proceedings of the Apollo 11 Lunar Science Conference, vol. 1*, pp. 965–988. New York: Pergammon Press.
Yoder CF (1981) The free librations of a dissipative Moon. *Philosophical Transactions of the Royal Society of London A* 303: 327–338.
Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
Yoder CF (1997) Venusian spin dynamics. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1087–1124. Tucson, AZ: University of Arizona Press.
Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303.
Yoder CF and Standish EM (1997) Measurements of the precession and rotation from Viking lander range data. *Journal of Geophysical Research* 102: 4065–4080.
Zeng L and Sasselov DD (2013) A detailed model grid for solid planets from 0.1 through 100 Earth masses. *Publications of the Astronomical Society of the Pacific* 125: 227–239.
Zharkov VN (1996) The internal structure of Mars: A key to understanding the origin of terrestrial planets. *Solar System Research* 30(6): 456–465.
Zharkov VN and Gudkova TV (1997) On the dissipation factor of Martian interiors. *Planetary and Space Science* 45: 401–407.
Zharkov VN and Gudkova TV (2000) Interior structure models, Fe/Si ratio and parameters of figure for Mars. *Physics of the Earth and Planetary Interiors* 117: 407–420.
Zharkov VN and Gudkova TV (2005) Construction of Martian interior model. *Solar System Research* 39: 343–373.

Zharkov VN and Gudkova TV (2009) The period and Q of the Chandler wobble of Mars. *Planetary and Space Science* 57: 288–295.

Zharkov VN, Gudkova TV, and Molodensky SM (2009) On models of Mars interior and amplitudes of forced nutations. 1. The effects of deviation on Mars from its equilibrium state on the flattening of the core-mantle boundary. *Physics of the Earth and Planetary Interiors* 172: 324–334.

Zharkov VN, Koshlyakov EM, and Marchenkov KI (1991) The composition, structure, and gravitational field of Mars. *Astronomicheskii Vestnik* 25(5): 515–547.

Zharkov VN, Molodensky SM, Brzeziński A, Groten E, and Varga P (1996) *The Earth and its Rotation. Low Frequency Geodynamics*. Heidelberg: Wichmann.

Zhong S and Zuber MT (2001) Degree-1 mantle convection and the crustal dichotomy of Mars. *Earth and Planetary Science Letters* 189: 75–84.

Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.

Zolotov MY, Sprague AL, Hauck SA II, Nittler LR, Solomon SC, and Weider SZ (2013) The redox state, FeO content, and origin of sulfur-rich magmas on Mercury. *Journal of Geophysical Research* 118(1): 118–146. http://dx.doi.org/10.1029/2012JE4274.

Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 220–227.

Zuber MT, Lemoine FG, Smith DE, Konopliv AS, Smrekar SE, and Asmar SW (2007) Mars Reconnaissance Orbiter radio science gravity investigation. *Journal of Geophysical Research* 112: E5. http://dx.doi.org/10.1029/2006JE002833.

Zuber MT, Smith DE, Lemoine FG, and Neumann GA (1994) The shape and internal structure of the Moon from the Clementine mission. *Science* 266: 1839–1843.

Zuber MT, Smith DE, Phillips RJ, et al. (2012) Topography of the northern hemisphere of Mercury from MESSENGER laser altimetry. *Science* 336: 217–220.

Zuber MT, Smith DE, Solomon SC, et al. (1992) The Mars Observer laser altimeter investigation. *Journal of Geophysical Research* 97: 7781–7797.

Zuber MT, Smith DE, Solomon SC, et al. (2008) Laser altimeter observations from MESSENGER's first Mercury flyby. *Science* 321: 77–79.

Zuber MT, Smith DE, Watkins MM, et al. (2013) Gravity field of the Moon from the Gravity Recovery and Interior Laboratory (GRAIL) mission. *Science* 339: 668–671.

Zuber MT, Solomon SC, Phillips RJ, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

10.03 Planetary Seismology

P Lognonné, Université Paris Diderot -Sorbonne Paris Cité, Institut de Physique du Globe de Paris, Paris, France
CL Johnson, University of British Columbia, Vancouver, BC, Canada; Planetary Science Institute, Tucson, AZ, USA

© 2015 Elsevier B.V. All rights reserved.

10.03.1	**Introduction**	65
10.03.2	**Lunar Results**	70
10.03.2.1	The Apollo PSE Data	70
10.03.2.2	Seismic-Velocity Structure: Crust and Mantle	72
10.03.2.3	Very Deep Interior	77
10.03.2.4	Mineralogical and Thermal Interpretation of Lunar Seismic Models	78
10.03.3	**Seismic Activity of the Moon and Terrestrial Planets**	81
10.03.3.1	Internal Seismic Activity	81
10.03.3.2	External Seismic activity: Artificial and Natural Impacts	86
10.03.4	**Atmospheric Seismology**	90
10.03.4.1	Theoretical Background	90
10.03.4.2	Mars Hum and Martian Atmospheric Sources	94
10.03.4.3	Venus Atmospheric Seismology	96
10.03.4.4	Giant Planets Seismology	99
10.03.5	**The New Step: Mars Seismology**	103
10.03.5.1	Interior Structure of Mars	103
10.03.5.2	Martian Seismic Noise	105
10.03.5.3	Body-Wave Detection	106
10.03.5.4	Normal Mode Excitation and Tidal Observations	109
10.03.5.5	Surface Waves	111
10.03.6	**Concluding Remarks**	113
Acknowledgments		114
References		114

10.03.1 Introduction

One hundred and twenty five years have passed since Von Rebeur-Paschwitz (1889) first detected a remote seismic event, and almost 70 years have passed since the creation of the first mean seismic models of the Earth based on body-wave travel times (Bullen, 1947). As a result, seismology is now generally accepted as the geophysical tool best able to determine the internal structure of a planet. Seismology has also led to a revolution in Earth science, especially since the advent of three-dimensional (3D) tomographic models of the Earth's mantle (e.g., Dziewonski et al., 1977; Woodhouse and Dziewonski, 1984) that depict major discontinuities in the mantle and reflect convection patterns and lateral variations in its temperature or mineralogy. For a complete and extensive description of seismology as applied to Earth, see volume 1 of *Treatise on Geophysics: Deep Earth Seismology*.

We are a long way from achieving this level of seismological knowledge for bodies other than Earth, however, and the seismic identification of the Earth's core, realized by R.D. Oldham one century ago (Oldham, 1906), remains the only example of detecting seismic waves refracted by a planetary core. The situation is similar for the detection of a planet's normal modes, with the first successful detection of normal modes on Earth following the great Chilean Earthquake in 1960 (e.g., Benioff et al., 1961). As a consequence (see the other chapters of this Treatise), we currently have no precise measurements of the core radius of the Moon or Mars; no direct information on the martian mantle structure, including its discontinuities; no definitive direct measurements of martian seismic activity; and no direct measurement of the mean crustal thickness of Mars.

The use of seismometers in planetary exploration was proposed early in the history of space missions (e.g., Press et al., 1960). Yet, almost 52 years after the launch of the first seismometer to a telluric body (Ranger 3, in 1962), planetary and small-body seismology has only been successfully studied on Earth's Moon. The Apollo program deployed a network of four seismic stations on the Moon (Latham et al., 1969, 1970a,b, 1971), as well as the short-lived Apollo 11 seismometer for passive monitoring, and three active seismic experiments were conducted during the missions of Apollo 14, 15, 16, and 17 (Kovach and Watkins, 1973a; Watkins and Kovach, 1972). The thirteen other extraterrestrial seismometer experiments that launched successfully never recorded any quakes (see **Table 1**). These nonyielding lunar missions include the seismometers (Lehner et al., 1962; Press et al., 1960) onboard the 3 Ranger lunar probes, which were lost with their missions in the early 1960s; the seismometer lost with the cancellation of the Apollo 13 Moon landing; and the gravimeter onboard the last Apollo 17 mission, which failed to operate despite expectations that it would extend the work of the previous Apollo instruments by providing long-period (LP) seismic data. Several attempts have been made to conduct seismic experiments on Mars. Two

Table 1 Summary and history of planetary seismology experiments delivered to launch

Mission	*Launch*	*Major mission events*	*Instrument description*	*Seismometer deployment*	*References*
Ranger 3	1962-01-26	Failure due to the booster. Moon missed	Vertical axis seismometer, with a free frequency of 1 Hz (mass: 3.36 kg)	Seismometer in a lunar capsule designed for a 130–160 km h^{-1} landing. Batteries powered for 30 days of operation	Lehner et al. (1962)
Ranger 4	1962-04-23	Failure of spacecraft central processor. Moon crash			
Ranger 5	1962-10-18	Failure in the spacecraft power system. Moon missed			
Apollo 11	1969-7-16	Successful installation. Powered by solar panel, worked during the first lunation and stopped after 21 days	Passive seismic experiment (PSE). Triaxis long-period (LP) seismometer and one vertical short-period (SP) seismometer, with resonance periods of 15 and 1 s, respectively (mass: 11.5 kg, power: 4.3–7.4 W)	Installation performed by crew. Seismometers were manually leveled and oriented with bubble level and sun compass. A sun protection/thermal shroud was covering the instruments. Power was delivered by a plutonium radioisotope thermoelectric generator for A12–14–15–16	Latham et al. (1969, 1970a,b)
Apollo 12	1969-11-14	Successful installation of a network of 4 stations. Except for the Apollo 12 SP seismometer and Apollo 14 vertical LP seismometer, all operated until the end of September 1977, when all were turned off after a command from the Earth. 26.18 active station years of data collected			
Apollo 14	1971-01-31				
Apollo 15	1971-07-26				
Apollo 16	1972-04-16				
Apollo 13	1970-4-11	Moon landing aborted. No installation of the PSE experiment, but the lunar crash of the Apollo 13 Saturn-IV upper stage was recorded by the A12 PSE			
Apollo 14	1971-01-31	Successful installation and operation of the active seismic experiments. Seismic sources were thumper devices (A14, A16) containing 21 (A14) and 19 (A16) small explosive sources and a rocket grenade launcher with three sources exploding up to 900 m on A16. Eight sources were used containing up to 2722 g of explosive and deployed up to about 2800 m by astronauts on A17	String of three geophones on A-14 and A-16 and on four geophones on A-17. Frequency was 3–250 Hz	Geophones were anchored into the surface by short spikes as they were unreeled from the thumper/geophone assembly	Watkins and Kovach (1972), Kovach and Watkins (1973a,b)
Apollo 16	1972-04-16				
Apollo 17	1972-12-07				
Apollo 17	1972-12-07	Deployment of the Lunar Surface Gravimeter. The gravimeter was unable to operate properly due to an error in the design of the proof mass	Gravimeter designed for gravity-wave detection. Additional LP vertical seismic output (10^{-11} lunar g resolution) for free oscillation detection, with a 16-Hz sampling	Installation performed by crew	Weber (1971), Tobias (1978)
Viking Lander 1	1975-08-20	Successful landing but instrument failure	SP instrument, with an undamped natural period of 0.25 s, a mass of 2.2 kg, a size of $12 \times 15 \times 12$ cm and a nominal power consumption of 3.5 W	The seismometer was installed on the Lander platform. No recentering was necessary, because the 3-axis seismometer had been designed to function even when tilted to up to 23°	Anderson et al. (1977a,b)
Viking Lander 2	1975-09-09	Successful landing and 19 months of nearly continuous operation. Too high wind sensitivity associated to the elastic recovery of the Viking			

(*Continued*)

Table 1 (Continued)

Mission	*Launch*	*Major mission events*	*Instrument description*	*Seismometer deployment*	*References*
		landing legs to the loading of the station by pressure fluctuations induced by the wind			
Phobos 1-2	1988-07-07 1988-07-12	Phobos 1 was lost during transfer to Mars and Phobos. Contact with Phobos 2 was lost just before the final phase of lander deployment, after Mars orbit insertion		Instrument onboard the long-service lander	Surkov (1990)
Mars 96 Small surface stations	1996-11-16	Failure of the block-D propulsion system in parking orbit. Earth re-entry. Two small stations and two penetrators lost	LP vertical-axis seismometer (0.1–4 Hz, 0.405 kg for the sensor) combined with a magnetometer. 55 mW of power	Seismometer in the small surface station. Semi-hard landing (200 g to 20 ms). Nominal operations for 1 martian year with nearly continuous operation with internal batteries for the first 90 days	Lognonné et al. (1998a)
Mars 96 Penetrators			High-frequency seismometer (10–100 Hz, 0.3 kg, 20 mW)	Seismometer in the penetrator. Hard landing. Nominal operations for 1 martian year	Khavroshkin and Tsyplakov (1996)
Rosetta	2004-03-04	Landing on the comet 67P/Churyumov-Gerasimenko planned a few months after rendezvous, expected on 2014-11-14	CASSE/SESAME experiment: three triaxial piezoelectric accelerometers and three piezoelectric active sources	Sources and accelerometers are housed within the landing gear's feet	Seidensticker et al. (2007)

Experiments that lead to the detection and interpretation of quake data are indicated by bolded names in the Mission column. The bolded and italicized name indicates a successful deployment without clear event detection or interpretation. The claimed detection of microseismic activity during the short-lived Venera lander mission might complete this list. All dates in the table are expressed as year-month-day.

Source: Ksanfomaliti LV, Zubkova VM, Morozov NA, and Petrova EV (1982) Microseisms at the Venera 13 and Venera 14 landing sites. *Soviet Astronomical Letters* 8: 241.

seismometers were deployed in 1976 by the Viking mission; however, one was never unlocked, and the other provided no convincing event detection after 19 months of nearly continuous operation (Anderson et al., 1977a,b). Twenty years later, both Optimism seismometers (Lognonné et al., 1998a) onboard small surface stations (Linkin et al., 1998) and the Kamerton short-period (SP) seismometers (Khavroshkin and Tsyplakov, 1996) onboard two penetrators were lost as part of the Mars 96 mission. Another 20 years will now pass before the return of the first data from the NASA InSight mission, which is planned to launch and land in 2016. Finally, for small bodies, the two SP seismometers onboard the Phobos landers (Surkov, 1990) never reached Phobos, and the SESAME/CASSE (Seidensticker et al., 2007) acoustic experiment, which is onboard the Rosetta lander en route to comet 67P/Churyumov–Gerasimenko, is now the only active experiment that might gather, in late 2014 early 2015, some seismic signals on a solid body other than Earth.

The seismometers developed for early missions used techniques that were pioneering, even compared to contemporary state-of-the-art Earth instruments. The Ranger seismometer was proposed by Press et al. (1960) and designed by the California Institute of Technology (Lehner et al., 1962) for the first missions to land on the Moon, and it was built to survive up to a 3000 g shock ($1\ g = 9.81\ m\ s^{-1}$). The Ranger device was a vertical-axis seismometer with a free frequency of 1 Hz and a mass of 3 kg, and it was one of the first digital instruments. Although the three Ranger probes failed, the technology was reused for operating one of the first terrestrial digital seismometers at Caltech (Miller, 1963). A new experiment was then proposed for the Surveyor mission, including a 3-axis LP seismometer and an SP vertical-axis seismometer, built at the Lamont Doherty Geological Observatory (Sutton and Latham, 1964), with a total mass of 11.5 kg. These models were later descoped to a single SP vertical-axis seismometer (Sutton and Steinbacher, 1967) before being canceled for the Surveyor program. As a result, this technology was not implemented until the Apollo missions to the Moon, and it was reused in terrestrial seismology, notably in the first ocean-bottom seismometers deployed in the 1970s. The deployment of seismometers was also planned for the Soviet Lunakhod, using a 1-Hz vertical fused quartz seismometer. This experiment was cancelled, however (Osika and Daragan, personal communication).

In July 1969, the Apollo 11 mission installed the first operational seismometer on a planetary body other than the Earth. The passive seismic experiment (PSE) consisted of a 3-axis LP seismometer with a resonance period of 15 s, as well as one vertical SP seismometer with a resonance period of 1 s. The total mass of both instruments and the electronic and thermal control module was 11.5 kg, and the power

consumption was between 4.3 and 7.4 W (Latham et al., 1969, 1970a,b). These seismometers were extremely sensitive. In flat-response mode, the LP seismometer was capable of detecting a displacement of 3×10^{-10} m at frequencies of 0.1–1 Hz, and in peaked-response mode, it could detect a displacement of 0.5×10^{-10} m at 0.45 Hz. For its part, the SP seismometer could detect a displacement of 0.5×10^{-10} m at 8 Hz. **Figure 1** shows the corresponding nominal acceleration response curves, as well as the noise recorded on the Moon. Practically, these instruments were unable to detect the continuous microseismic noise of the Moon, instead recording only the significant background noise associated with thermal effects in the shallow subsurface around sunrise and sunset. The Moon noise recorded on the vertical-component seismometer was as low as 10^{-10} ms^{-2} Hz$^{-1/2}$ at frequencies in the range 0.1–1 Hz (Earth's microseismic peak). This level of sensitivity allowed the detection of about 13 050 seismic signals between July 1969 and September 1977 (a mean of about four quakes per day), and many more unreported events were recorded by the SP instruments. The most recent LP event catalog is available online at the University of Texas, at ftp://ftp.ig.utexas.edu/pub/PSE/catsrepts/ (Nakamura et al., 2008).

The first PSE experiment was powered by solar panels, and it survived the first lunar night, but ultimately failed due to high temperatures during the second lunar day (Latham et al., 1969, 1970a,b). Subsequent seismometers used a radioisotope thermoelectric generator that supported continuous day and night operation. A seismic network of four stations (**Figure 2**) was installed by Apollo 12, 14, 15, and 16, and, as discussed in multiple lunar and seismological reviews, this network has been critical for understanding the internal structure of the Moon (e.g., Hood, 1986; Hood and Zuber, 2000; Lognonné,

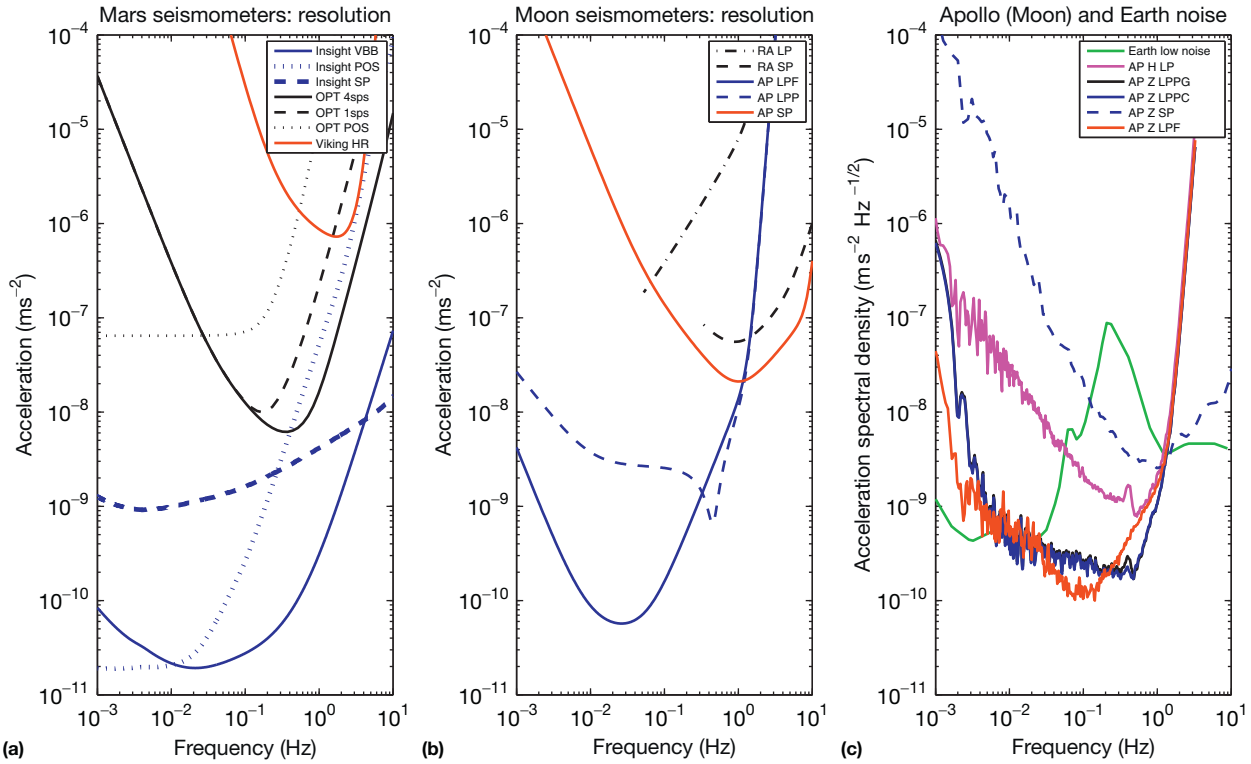


Figure 1 Response curves of the past Mars (a, left) and Moon (b, middle) seismometers. (a) The resolutions of the Viking seismometer in the high data rate mode (Viking HR) and the Optimism seismometer in the long period position mode (OPT POS), the velocity mode at 1 sample per second (OPT 1sps), and the velocity mode at 4 samples per second (OPT 4sps). The figure also shows the expected resolution of the InSight seismometers, which is defined here as the RMS in one octave bandwidth. The blue continuous and blue dashed curves are the velocity flat and acceleration flat outputs of the very broad band instrument (InSight VBB and InSight POS), while the long dashed blue one is the velocity flat output of the SP instrument (InSight SP). VBB components are sensitive to acceleration along an oblique about 30° with respect to horizontal, while the SP components are either vertical or horizontal. At 1 Hz, the expected resolution of the InSight VBB sensor is expected to be 10^3 better than Viking. (b) The Moon, with the long-period (RA LP) and short-period (RA SP) analog outputs produced by the Ranger seismometer, the flat mode (AP LPF) and peaked mode (AP LPP) of the Apollo LP seismometer, and the SP seismometer (AP SP). The transfer functions shown in the figure are generic. (c) Different sources and levels of seismic noise on the Moon. The noise levels recorded on the Moon by the different channels of the Apollo seismometers are compared with the Earth-based low noise model of Pedersen (1993). The noise levels are from the Apollo 12 instrument for the vertical LP flat component (AP Z LPF) and from the Apollo 14 instrument for the vertical SP component (AP Z SP). These noise levels likely represent an upper estimate of the Moon noise and may be related to the instruments. For the other components, note that the peaks near 0.45 Hz of the peaked vertical records (AP Z LPPG) and peaked horizontal records (AP H LP) might be related to differences between the Apollo 12 transfer function at the time of the recording and the preflight generic transfer function. The noise corresponding to the curve AP Z LPPC is obtained by changing the feedback parameters for the transfer function of the instrument in order to minimize the peak in the noise PSD. Changes are on the order of 10% for the parameters K2 and h. See the Apollo transfer function at http://darts.jaxa.jp/planet/seismology/apollo/The_Apollo_Seismometer_Responses.pdf).

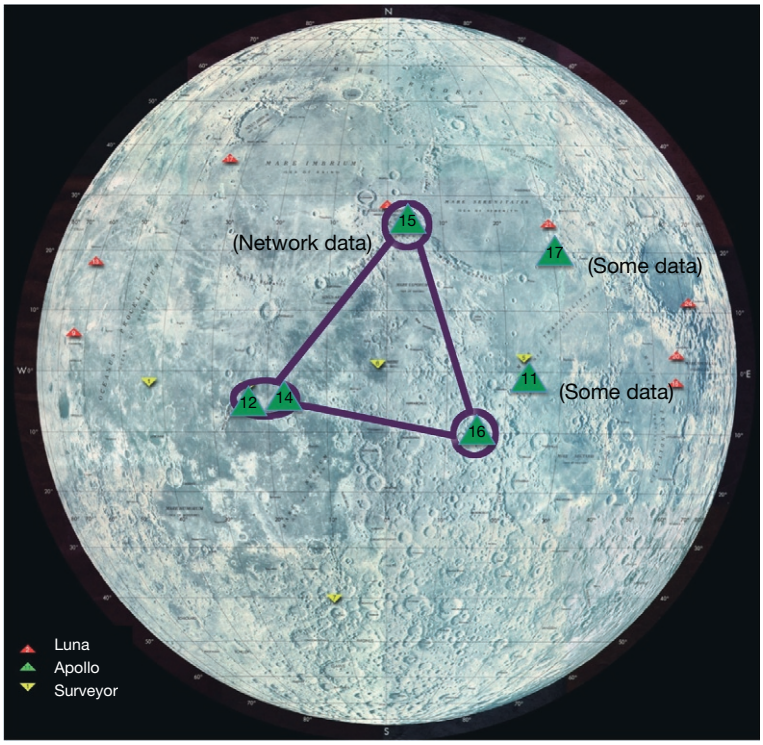


Figure 2 Configuration of the Apollo seismic network, with each Apollo seismometer represented by a green triangle. A12 is in the Oceanus Procellarum area at 3.01° S, 23.42° W. A14 is located near the crater Fra Mauro at 3.64° S, 17.47° W. A15 is at the foot of the Apennine Mountains at 26.13° N, 3.63° E, and A16 is just north of the crater Dolland at 8.97° S, 15.50° E. Note that all stations were on the near-side, making core seismic studies almost impossible. Additional seismic data have also been recorded at the Apollo 11 and Apollo 17 sites. The soviet Luna and early US Surveyor landing sites are given for completeness. Base map credit: NASA.

2005). Given its low seismic noise, the Moon has also been proposed as a good place for detecting astrophysical gravitational waves (e.g., Gusev and Kurlachev, 1976; Weber, 1960), and a gravimeter designed for that purpose was installed onboard Apollo 17. This gravimeter failed to operate properly, however, due to a design failure. More recently, the possibility of strange quark matter detection (Banerdt et al., 2006) has been proposed, reigniting interest in the Moon as a site for the placement of seismic or gravity sensors for astrophysics research.

In this chapter, we summarize the experimental and seismological data in **Section 10.03.2**. We devote some attention to 1D seismic-velocity models for the Moon's crust and mantle, and we discuss differences among current models, especially with regard to their likely causes and implications. We also explore the current understanding of the Moon's mineralogical structure derived by integrating the Apollo seismic data with other geophysical and petrological constraints, and, in **Section 10.03.3**, we discuss the seismic activity of the Moon in the context of other terrestrial planets.

We address the field of atmospheric seismology in **Section 10.03.4**, including the basic theory behind the coupling of solid-body modes to the atmosphere and ionosphere. We also discuss the giant planets, especially Jupiter, given that, for about 20 years, Earth stations have monitored Jupiter's atmospheric signals associated with continuously excited global oscillations (Lognonné and Mosser, 1993; Mosser, 1995). We then develop a comparative study of atmosphere-interior seismic coupling on Mars and Venus and discuss the prospect of remotely sensed seismology.

Starting with the 1976 deployment of the Viking seismometers, the seismic exploration of Mars has been much less successful than the exploration of the Moon (Anderson et al., 1977a,b). During the initial mission, only the Viking 2 seismometer worked, as the seismometer on Viking 1 lander failed to unlock. The sensitivity of the Viking 2 seismometer was one order of magnitude less than the SP Apollo seismometer for periods shorter than 1 s, and five orders less than the LP seismometer for periods longer than 10 s (see **Figure 1**). No events were convincingly detected during the seismometer's 19 months of nearly continuous operation, and, as shown by Goins and Lazarewicz (1979), this absence of recorded events was probably related to the inadequate sensitivity of the seismometer in the frequency bandwidth of teleseismic body waves, as well as the device's high sensitivity to wind noise (Nakamura and Anderson, 1979). In 1996, the Mars 96 mission was launched, with each of the mission's two small stations (Linkin et al., 1998) carrying an OPTIMISM seismometer in its payload (Lognonné et al., 1998a). The sensitivity of the OPTIMISM seismometer was improved by about two orders of magnitude relative to the Viking seismometers at frequencies of 0.5 Hz, and as a result, the OPTIMISM devices were better adapted to teleseismic body-wave detection. The Mars 96 mission was lost shortly after its launch, however, and the seismic

exploration of Mars remains an outstanding avenue of investigation in solar system science. Thus, determining the seismicity of Mars is the main goal of the InSight NASA mission, which will deploy the first geophysical observatory on the red planet. Two Apollo-class 3-axis seismometers will be deployed by InSight, for both LP and SP detection. The expected performances of these devices are shown in **Figure 1**. Compared to the Apollo seismometers, the InSight seismometers have slightly better resolution, and they will likely be solely limited by the martian environment and possibly microseismic noise. We will return to the discussion of Mars seismology and InSight in **Section 10.03.5**, as it leads the way for future surface-based exploration of other bodies. We will then discuss the current understanding of the martian interior, along with noise levels and other issues relevant to the InSight strategy for marsquake detection.

10.03.2 Lunar Results

10.03.2.1 The Apollo PSE Data

From 1969 to 1977, the Apollo Lunar Surface Experiment Package (ALSEP) recorded the lunar PSE data that now provide a unique and valuable resource for determining the interior structure of the Moon. The Apollo seismic 'network' comprised four stations at Apollo sites 12, 14, 15, and 16. Stations 12 and 14 were about 180 km apart, and, together, they formed one corner of an approximately equilateral triangle, with stations 15 and 16 at the other corners, about 1100 km apart (**Figure 2**). The lunar event catalog (Nakamura et al., 2008) documents 13 058 events recorded by the network. Recorded events exhibit different signal characteristics and originate from five types of sources originally classified as artificial impacts (9), meteoroid impacts (~1700), shallow moonquakes (28), deep moonquakes (~1360), and unclassified events. Many of the originally unclassified events have since been classified as deep moonquakes, with a total of 7084 deep moonquakes identified (as reported in Nakamura et al., 2008).

Compared with terrestrial quakes, moonquakes are small-seismic-moment events, and, as a result, the moonquakes recorded by the Apollo missions generated small signals, often with amplitudes smaller than the microseismic noise on Earth, even in Earth's best seismic vaults. At first, events were visually identified in the lunar data, using hardcopy printouts and overlays of the seismograms recorded at Apollo stations 12, 14, 15, and 16. The criteria used to identify and classify seismic sources included the time interval from the start of a signal to its maximum amplitude (rise time), the dominance of low-frequency versus high-frequency content in the records, and the presence of compressional (P) and shear (S) waves (see review in Lammlein, 1977 and details in Lammlein, 1973 and Lammlein et al., 1974). Shallow events are also referred to in the literature as high-frequency teleseismic events, because of their unusually high frequency content and poor depth location (Nakamura, 1977a, 1980; Nakamura et al., 1974a, 1979). Analyses of data during early ALSEP operations suggested that deep moonquakes have two important characteristics (Nakamura et al., 1980) that were subsequently used to help identify such events. These characteristics are similar waveforms and tidal periodicities in moonquake occurrence times (Ewing et al., 1971; Lammlein, 1977). The main properties of the seismic sources, including seismic moment, stress drop, and rupture times, are given in **Section 10.03.3**. **Figure 3** also shows examples of waveforms recorded at the Apollo stations, according to source type. In general, events become apparent in the seismic readings only after data filtering. Filtering is required because the raw waveform data are dominated by large amplitude signals that result from the expansion and contraction of the thermal Mylar shroud at sunrise and sunset (Duennebier and Sutton, 1974; see Figure 3 of Bulow et al., 2005).

During the Apollo era, digital data were extracted from the complete, continuous PSE records for time windows encompassing identified events, and the digital data were then stored as the 'event' waveform data set. This derived data set and the original continuous records are now available, with documentation, from the IRIS Data Management Center (http://www.iris.edu/dms/nodes/dmc/) and at the ISAS DARTS web site (http://www.darts.isas.jaxa.jp/planet/seismology/apollo/index.html, Nagihara et al., 2011). Recent analyses of the Apollo 17 gravimeter complement the lunar data set with seismic records from the Apollo 17 landing site (Kawamura et al., 2010a,b), including records for deep moonquakes. Although the generic transfer functions of the Apollo seismometer have been published (see, e.g., on ISAS DARTS) and are generally used in most publications, it is possible that aging led to some change in these transfer functions. For example, when corrected with the generic transfer function, the noise spectrum of the Apollo 12 vertical records in peaked mode shows significant peaks near the peaked frequency of about 0.45 Hz (blue curve of the right box of **Figure 1**). These peaks decrease significantly, however, when slight changes are made to some of the transfer function parameters, as when the coil-magnet efficiency and the damping constant of the natural suspension are changed by −6% and −16%, respectively, from their nominal values for the light blue curve. For this reason, time-related changes in seismometer function will be kept in mind when detailed analyses are done on Apollo seismic data, including on the subsurface resonances. Although accurate time-dependent transfer functions can be obtained by processing calibration signals, which were recorded periodically (usually once a week) throughout the experiment, the coefficients of these functions (i.e., gain, pole, and zero) for each of the Apollo instruments throughout the 7 years of operation are not available in the literature.

Later work (Nakamura, 2003) has classified a significant fraction of identified, but previously unclassified, events (Nakamura et al., 1981) and has led to a nearly fivefold increase in the number of cataloged deep moonquakes. Additionally, work by Bulow et al. (2005) showed that the waveform repeatability of deep moonquakes can be exploited to search for additional unidentified events associated with known deep source regions. Implementation of this algorithm led to an approximately 35% increase in the number of events at the nine most seismically active (at the detection level of the Apollo seismometers) source regions (Bulow et al., 2005, 2007). Despite the seemingly large number of lunar seismic events, the poor signal-to-noise ratio for emergent arrivals in most records means that it is only possible to accurately measure arrival times and locate events for a subset of the lunar catalog

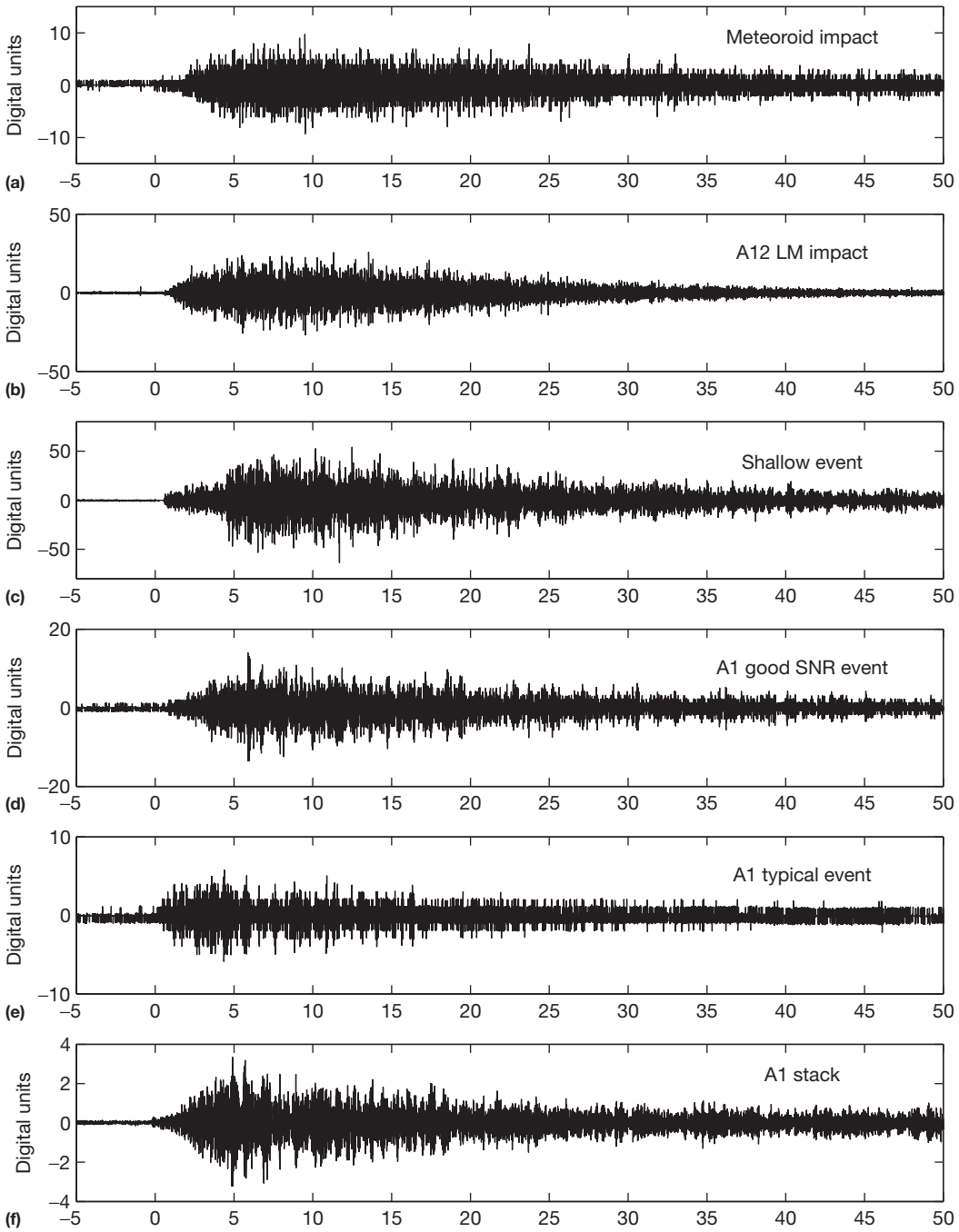


Figure 3 From top to bottom, examples of waveforms for natural impacts (a), artificial impacts (b), a shallow event (c), and deep moonquakes (d–f) recorded on LP channels by the Apollo passive seismic network. All records span 55 min, with the *x*-axis representing time in minutes and the *y*-axis representing digital units. The following characteristics aided the identification of different seismic sources in the Apollo data: (1) The rise time (defined as the time from event onset to maximum signal amplitude) of moonquakes is typically less than that of meteoroids – the rise time in (c), (d), and (e) is approximately 5 min, while the rise time in (a) is approximately 10 min; (2) Deep moonquakes typically have low amplitude and produce poor data quality, as in (e). However, the waveform repeatability of deep moonquakes from distinct source regions allows the use of stacking to improve the signal-to-noise ratio, as in (f).

data (Figure 4). Small event magnitudes and restricted receiver locations mean that events for which travel times can be clearly identified mainly occur on the near-side of the Moon.

The Apollo passive seismic data set has contributed enormously to our understanding of the internal structure of the Moon. The temporal and spatial distributions of lunar seismicity provide constraints on subsurface structure. Of particular importance is the evidence of brittle failure at depths of 700–1200 km, with failures apparently concentrated in the 800–1000-km depth interval. Unfortunately, the small

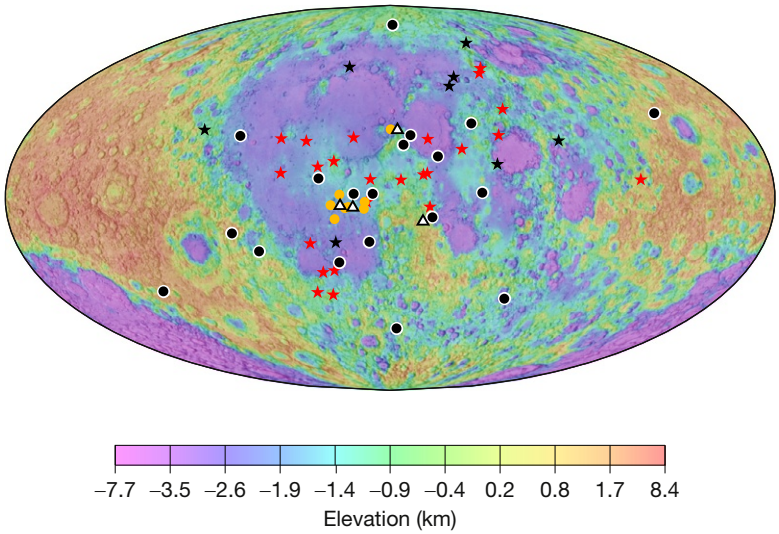


Figure 4 Seismic events located by Lognonné et al. (2003) and used in recent velocity models (Gagnepain-Beyneix et al., 2006; Lognonné et al., 2003). The figure shows Apollo stations 12, 14, 15, and 16 (white triangles), along with the locations of 24 deep-event clusters (red stars), 8 shallow events (black stars), 19 meteoroid impacts (black circles), and 8 artificial impacts (orange circles). The background color map is the Clementine topography (Aitoff equal area projection), shaded using a global digital image map (USGS map I-2769, 2002).

magnitude of these events, combined with the limited number and geographical distribution of receivers, and associated errors in the locations of deep moonquakes (see, e.g., Hempel et al., 2012 for a recent analysis of the location errors), means that it has not been possible to unambiguously establish a focal mechanism for any of the deep moonquakes (Koyama and Nakamura, 1980). Seismic-velocity models for the crust and mantle (see **Section 10.03.2.2**) constrain the physical properties of the lunar interior (including temperature) and provide indirect constraints on mineralogy and the thermal evolution of the Moon. Studies of seismic attenuation indicate a dry lunar mantle (Nakamura and Koyama, 1982). Attempts have been made to investigate lunar normal modes (Khan and Mosegaard, 2001) to constrain the core size and state. However, these have largely been unsuccessful because of the limited bandwidth of the Apollo seismometers (**Figure 1**).

The Moon's shallow structure has received less attention in the literature, but it is of particular interest given the results from the Gravity Recovery and Interior Laboratory (GRAIL) mission (see below). The long coda in the observed waveforms likely result from the combined effects of low attenuation and a highly scattering near-surface layer with low seismic velocity (Latham et al., 1970a; Nakamura, 1976, 1977b). These coda can thus be used to probe the structure of the brecciated upper crustal layer, or megaregolith. Early lunar studies investigated the geological structure in the immediate vicinity of the receivers. For example, Nakamura et al. (1975) estimated regolith thicknesses of a few meters beneath Apollo stations 11, 12, and 15, with an underlying layer displaying P-wave velocities of 250–400 m s^{-1}, and Horvath et al. (1980) calculated seismic-velocity profiles down to a depth of about 200 m. The coda decay can be characterized by the quality factor Q_c (where Q_c^{-1} represents the fraction of energy dissipated per cycle at a given frequency), and, for a lunar module impact at Apollo station 12, Latham et al. (1970a) found $Q_c \sim 3600$, a value much larger than those recorded for terrestrial crustal rocks. Blanchette-Guertin et al. (2012) have undertaken the first systematic survey of the coda characteristics of impacts and moonquakes recorded at all four Apollo passive seismic stations. The results indicate decay times and Q_c increase with depth, and Q_c increases with increasing frequency and epicentral distance for shallow events. The results are consistent with scattering in a global, low-velocity and low-attenuation megaregolith layer with a size distribution of scatterers (blocks) that have more small-scale scatterers than large-scale scatterers. In addition, this study finds that scattering can be intensified locally in one or more frequency bands due to local structure.

In the following section, we discuss attempts to establish seismic velocity profiles for the lunar crust and mantle, and we also highlight GRAIL mission results related to lunar crustal structure. Because of their potential to provide significant insights concerning the lunar interior, these seismic-velocity models have been used for many seismological, mineralogical, and thermal studies of the Moon.

10.03.2.2 Seismic-Velocity Structure: Crust and Mantle

Due to scattering in the lunar seismograms, as well as the limited computational resources of the 1970s, most investigations of the lunar interior have used the arrival times of body-wave phases to investigate plausible 1D models for seismic-velocity structure. This approach, completed using normal modes and surface-wave data, has been applied with great success in terrestrial seismology through a series of efforts leading to the long-standing preliminary reference Earth model (PREM) for Earth's interior structure (Dziewonski and Anderson, 1981) and to the body-wave IASP91 model (Kennett and Engdahl, 1991). Lunar studies use the arrival times of direct P and S waves, picked from seismograms for as many channels and stations as possible. Individual seismograms are used for impacts and shallow moonquakes. For deep moonquake sources, all records of a specific source region are stacked to improve the

signal-to-noise ratio (Lognonné et al., 2003; Nakamura, 1983). Source-to-receiver P- and S-wave travel times depend on the receiver location (known), the source location (generally unknown, though natural impacts occur on the surface, and artificial impact locations are known), the source time (unknown, except for artificial impacts), and the interior velocity structure (unknown). The standard seismological approach involves simultaneously solving for source locations and times, along with interior velocity structure (e.g., Shearer, 1999).

As measured from the lunar seismograms, P- and S-wave arrival times have been either inverted or forward-modeled to estimate lunar seismic-velocity profiles (e.g., Gagnepain-Beyneix et al., 2006; Garcia et al., 2011; Goins et al., 1981b; Khan and Mosegaard, 2002; Khan et al., 2000, 2007; Lognonné et al., 2003; Nakamura, 1983; Nakamura et al., 1976). As the construction of lunar seismic-velocity models has been reviewed elsewhere (Lognonné, 2005), we do not repeat those reviews here. Instead, we restrict our discussion to five 1D models for P- and S-wave velocity based on travel-time data (Gagnepain-Beyneix et al., 2006; Garcia et al., 2011; Goins et al., 1981b; Khan et al., 2007; Nakamura, 1983) and one model derived solely from mineralogical modeling without a priori seismic constraints (Kuskov et al., 2002). Regardless of their forms, all discussed models illuminate the main unresolved questions about lunar interior structure (**Figure 5**).

Figure 5 shows the P- and S-velocity models of Goins et al. (1981b), Nakamura (1983), Khan et al. (2007) (see also Khan and Mosegaard, 2002; Khan et al., 2000), Gagnepain-Beyneix et al. (2006), and Garcia et al. (2011). We have chosen these studies because they are all primarily based on the Apollo seismic data, and they exemplify the state-of-the-art calculations made immediately after the publication of the complete lunar event catalog (Goins et al., 1981b; Nakamura, 1983), the alternative modern probabilistic techniques applied to the early travel-time data set (Khan et al., 2007), recent revised travel-time picks and derived velocity models (Gagnepain-Beyneix et al., 2006; Lognonné et al., 2003), and additional constraints on the deep moon structure based on transversely polarized, reflected S waves (Garcia et al., 2011).

The model of Nakamura (1983) (N83) used travel times from a total of 81 events, and it specified four crustal layers, with mantle interfaces at 270 and 500 km depth. The model minimized the root-mean-square misfit of the predicted and observed travel times, linearizing about the starting model. Nakamura (1983) stressed that the choice of interface depths and the specification of constant velocity layers was arbitrary. P- and S-wave travel times were repicked for 59 events by Lognonné et al. (2003) and used in models by Lognonné et al. (2003) and Gagnepain-Beyneix et al. (2006) (L03 and GB06, respectively) that specify three to four mantle layers and four to six crustal layers. The Khan et al. (2007) model used a Monte Carlo simulation approach, combined with Bayesian inference, to find families of compositional, density, and thermal models compatible with the L03 travel-time data set

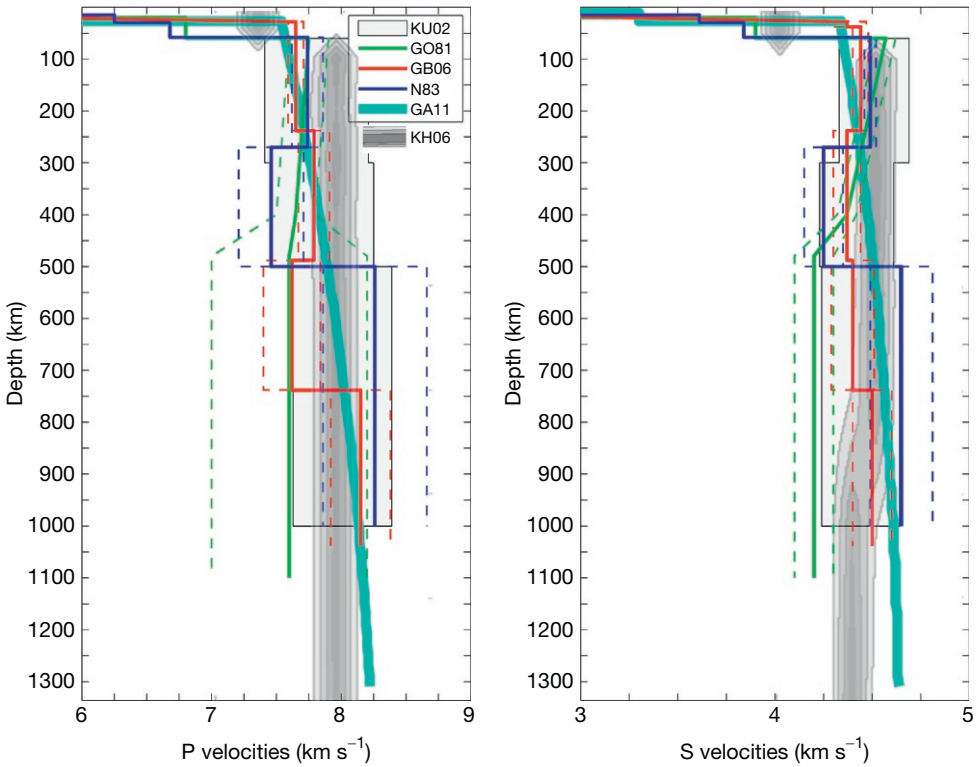


Figure 5 P- and S-wave velocity models versus depth: green – Goins et al. (1981b) (G081), red – Gagnepain-Beyneix et al. (2006) (GB06), dark blue – Nakamura (1983) (N83), light blue – Garcia et al. (2011) (GA11). From light to dark, the gray zones are the zones of increasing probability from Khan et al. (2006) (KH06). The light gray represents the petrology–geochemistry model of Kuskov et al. (2002) (KU02). Values for mantle velocities and associated 1σ uncertainties for each model can be found in Table 1 of Lognonné (2005) and are plotted using dashed lines for the three first models.

and the lunar mass and moment of inertia. A similar inversion approach was used by Kuskov et al. (2002), but only the lunar mass and moment of inertia were inverted, and thus, the resulting seismic-velocity models could be compared in a predictive sense with those independently derived from the travel-time data. The most recent, comprehensive lunar seismic-velocity model is the model created by Garcia et al. (2011) (GA11). This study uses the travel-time data set of L03. GA11 takes the crustal velocity structure from GB06 and constructs a series of mantle velocity models for different core radii. For each core radius, core densities are computed that match the mass, moment of inertia, and Love numbers. GA11 then uses horizontally polarized, reflected S waves, combined with previously published attenuation parameters (Nakamura and Koyama, 1982; Nakamura et al., 1982), to establish the preferred core radius (see also **Section 10.03.2.3**) and, hence, the preferred overlying mantle-velocity model.

Crustal structure is primarily controlled by impacts. Seismic ray theory predicts that near-receiver impacts (mostly the artificial impacts) will have rays that turn within the crust. In addition, distant impacts produce seismic energy that travels through the crust near the source location, and then again near the receiver location. Thus, in theory, the travel times of P and S waves should constrain crustal thickness and velocity structure, with the strongest constraints near the longest-lived Apollo stations (12 and 14). Additional data for investigating crustal structure come from observations of S-to-P converted phases (Vinnik et al., 2001), the data for which have been recorded at stations A12 and A16 (at A14 and A15, data were either missing or too noisy). Converted phases were only detected at the Apollo 12 station and originated from a seismic-velocity discontinuity interpreted as the crust–mantle boundary. The terrestrial seismological literature refers to such studies as receiver function studies. The combination of the differential travel time and the relative amplitude of the primary and converted phases provides information on the depth of and the impedance contrast across the discontinuity. However, the lunar data set is less than ideal. More recent studies use 17 (L03) or 14 (GB06) travel times from 8 (L03) or 7 (GB03) artificial impacts and 88 travel times from 19 natural impacts. As the latter require estimates of three (latitude, longitude, and source time) source parameters, the total number of degrees of freedom is reduced to 48 (L03) or 45 (GB06). Travel times and relative amplitudes for converted phases from only one receiver location are sufficiently reliable for incorporation into velocity-profile modeling (L03, GB06).

Figure 5 shows the significant variability among crustal velocity models, particularly regarding crustal thickness. By analogy with Earth, the lunar seismological crust-mantle interface is defined by a velocity discontinuity or steep gradient (>0.1 km s^{-1} km^{-1}), below which the P-wave velocity should attain a value of at least 7.6 km s^{-1}. Specifically, depending on the model, crustal thickness is estimated to be 58 km (Goins et al., 1981b; *N83*, and see also earlier work by Toksöz, 1974), ~38–45 km (Khan and Mosegaard, 2002; KM07; Khan et al., 2000), and 30 km (GB06) (Figure 5). The velocity structure at depths of 30–60 km is quite sensitive to a few travel-time data (see discussion in GB06). Figure 6 shows the sampling of the lunar interior by way of ray paths connecting the events and receivers shown in Figure 4.

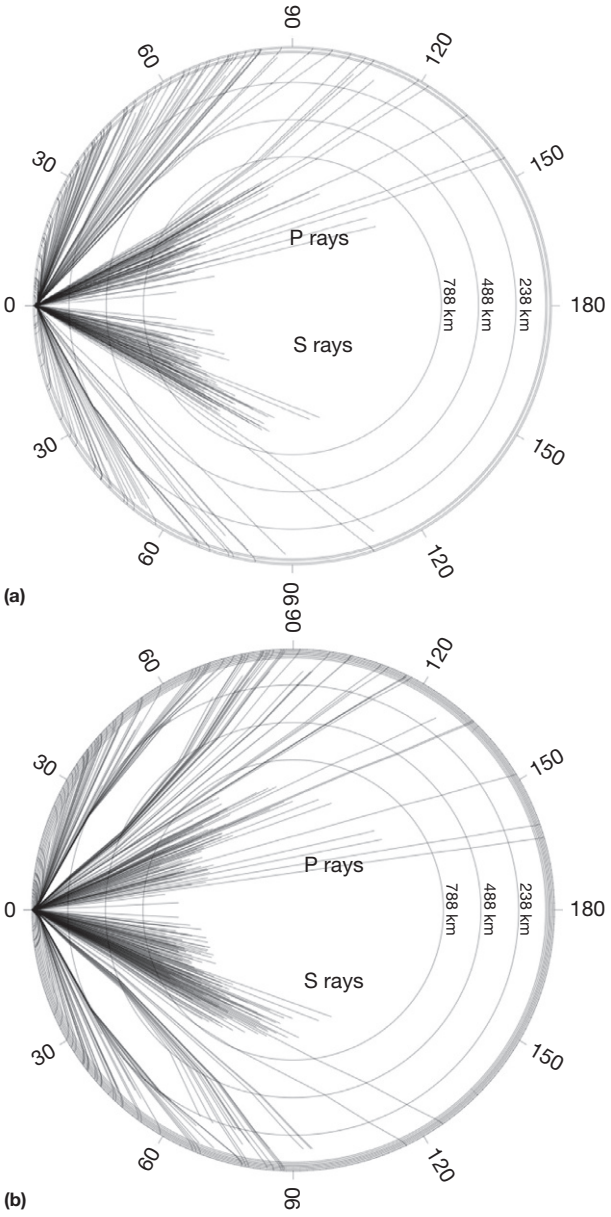


Figure 6 Summary of the source–receiver seismic ray paths for two different seismic models of the Moon. The top model (a) is determined using the travel times of Gagnepain-Beyneix et al. (2006), and the bottom model (b) is determined using the travel times of Nakamura (1983). Rays are plotted with the stations all located at 0° and the source at the actual depth and epicentral distance from the station. Note the rays associated with deep moonquakes. The crustal thickness in (a) is 30 km, below which there is a 10-km transition zone with a shear wave gradient. The crustal thickness for model (b) is 58 km. Despite differences in mantle sampling, no rays sound the lunar core for both models. Reprinted from Gagnepain-Beyneix J, Lognonné P, Chenet H, and Spohn T (2006) Seismic models of the moon and their constraints on the mantle temperature and mineralogy. *Physics of the Earth and Planetary Interiors* 159: 140–166, with permission from Physics of the Earth and Planetary Interior.

The ray path coverage is uneven with changes in depth, and, while hard to see in Figure 6, the sampling of very shallow depths (<60 km) is poor. While uncertainties in crustal

thickness remain, four main conclusions can be drawn from the seismological data. First, although the profiles shown in Figure 5 are 1D velocity profiles, the source–receiver geometries and ray theory indicate that the averaged crustal thickness primarily reflects the crustal structure at the Apollo 12 and 14 sites. Second, several recent tests indicate that the crust at the Apollo 12/14 sites is more likely to have a thickness of 30–45 km (KM02, L03, GB06), rather than the ~60-km thickness previously estimated (N83). Third, although not discussed here, an attempt has been made to investigate geographical variations in crustal thickness, using the Apollo seismic data (Chenet et al., 2006). While only a few estimates are possible, they show a high level of agreement with relative crustal thickness variations deduced from gravity and topography data prior to the GRAIL mission. Finally, high-resolution gravity and topography data from the GRAIL (Zuber et al., 2013) and Lunar Reconnaissance Orbiter missions have shown that the bulk density of the lunar highlands is 2550 kg m^{-3}, which is much lower than typically assumed (Wieczorek et al., 2013), but the new data also produces estimates of crustal thickness similar to the seismically derived crustal thicknesses of 30 km (L03, GB06) to 38 km (Khan and Mosegaard, 2002).

Estimates of seismic-velocity structure within the crust are variable (Figure 5). While literature has addressed the range of estimates, the variability among models mainly reflects the limited data set. Perhaps the most satisfactory summary of the seismic data asserts that the recorded seismic activity is broadly consistent with a crustal structure in which there are two major compositional layers – an upper anorthositic (lower seismic velocity) and lower noritic (higher seismic velocity) crust – consistent with inferences from gravity and topography data (Wieczorek et al., 2006). Seismologically, higher velocities are associated with the noritic lower crust. Some suggestions of a midcrustal reflector have been made (e.g., the 20-km discontinuity of Khan and Mosegaard, 2002). Furthermore, as discussed in Section 10.03.2.1, the megaregolith likely results in scattering of seismic energy and, thus, significantly reduced seismic velocities in the upper ~1 km (GB06 and see review in Wieczorek et al., 2006). This scattering is supported by the high average crustal porosities inferred from GRAIL (Wieczorek et al., 2013), and, in fact, a layer with v_p ~1 km s^{-1} may extend to over 3 km in depth, at least beneath the Apollo 17 site (Nakamura, 2011). This high near-surface porosity contributed to (and can explain) earlier thicker estimates of the lunar crustal thickness.

Despite the variability in the shallow structure, upper mantle (<300 km) velocity estimates are generally in agreement at around 7.7 km s^{-1} for P waves and 4.5 km s^{-1} for S waves (L03, N83, GB06). Velocity estimates in the 270–1000-km region are constrained primarily by travel-time picks from deep events. A midmantle (300–500 km depth) low-velocity zone or negative velocity gradient was suggested by N83, and this low-velocity zone is permitted, although not required (Figures 5–7 and following discussion). In particular, the crustal structure of N83 results in very poor ray coverage of the depth region from 300 to 500 km (Figure 6). The trade-off is that improved coverage of this region can be obtained (e.g., with thinner crustal thicknesses) at the expense of poor coverage at 500–800 km of depth (Figure 6, and see the discussion in GB06). The smoothly varying P- and S-wave mantle-velocity profiles of G11 lie between all previously proposed 1D models. Seismic velocities below ~1000 km are essentially unconstrained (Figure 6), although the absence of moonquakes below this depth may be consistent with possible high shear-wave attenuation, the latter having been proposed to explain the lack of shear-wave arrivals from the few detected far-side deep moonquakes (Nakamura et al., 1973). Inversions of the seismic data for a lunar-temperature profile (Gagnepain-Beyneix et al., 2006) suggest temperatures of about 1200 °C at this depth. All these observations argue for a warmer, more ductile lower mantle.

The effect on seismic-velocity-model resolution of ray-path coverage is also illustrated by a sensitivity analysis of GB06. We examine the sensitivity of the travel-time data set to perturbations to GB06. Weighted root mean square (RMS) misfits are calculated for velocity models in which the velocity is as specified in GB06, except in one depth interval. In this depth interval, the velocity is either increased (red line, Figure 7) or decreased (blue line, Figure 7) by 10% relative to the published value. Depth intervals are 100 km thick, except where the published model comprises a thinner layer. In the case of a thinner layer, the published layer thickness is used. The RMS misfit between the travel times predicted by GB06 (Figure 5) and the measured travel times is ~5.9 s, which is comparable to the RMS of the errors assigned to the data (~6.2 s). We instead show a weighted RMS misfit (Figure 7) for which each travel-time residual (observed minus predicted) is divided by the uncertainty in the travel-time pick (1, 3, or 10 s, see Lognonné et al., 2003), and the total RMS is further normalized such that GB06 fits the data to a weighted RMS of 1.0.

We then investigate perturbations to the P-wave and S-wave velocity models, calculating the RMS misfit using only the P- or S-wave arrival times, respectively (Figure 7(b) and 7(c)). In each of these cases, the v_p/v_s ratio can vary relative to the ratio specified in GB06. The travel-time data are more sensitive to perturbations in v_s than to perturbations in v_p, with the corresponding RMS misfits showing greater variation (Figure 7(c)). Models with a v_p/v_s ratio lower than the ratio from GB06 (higher v_s, Figure 7(c), or lower v_p, Figure 7(b)) are acceptable when compared with GB06, however, and models with lower ratios can even result in an improved misfit to the data set. (The exception is the depth interval 1000–1100 km, which contains the source depths of some deep moonquakes.) This may seem surprising at first, but one should remember that the analysis of GB06 is an inverse approach with a priori constraints applied. In other words, the best-fit model is determined according to some a priori criteria not applied here. Models with higher v_p/v_s ratios (i.e., increased v_p or decreased v_s) have larger RMS misfits than GB06, and with the exception of the perturbations to v_p in the depth range 500–800 km, these models result in more than a 10% RMS misfit increase, with the increase being as high as 50% for −10% perturbations to v_s in the depth range 100–200 km.

Somewhat different results are obtained if v_s and v_p are both allowed to vary, holding the v_p/v_s ratio constant. At depths shallower than 200 km, the travel-time observations allow for discrimination between plausible models. In particular, decreases in seismic velocity at depths <200 km generate increases in misfit levels of more than 30%. At depths of 300–500 km, there is some sensitivity to seismic structure,

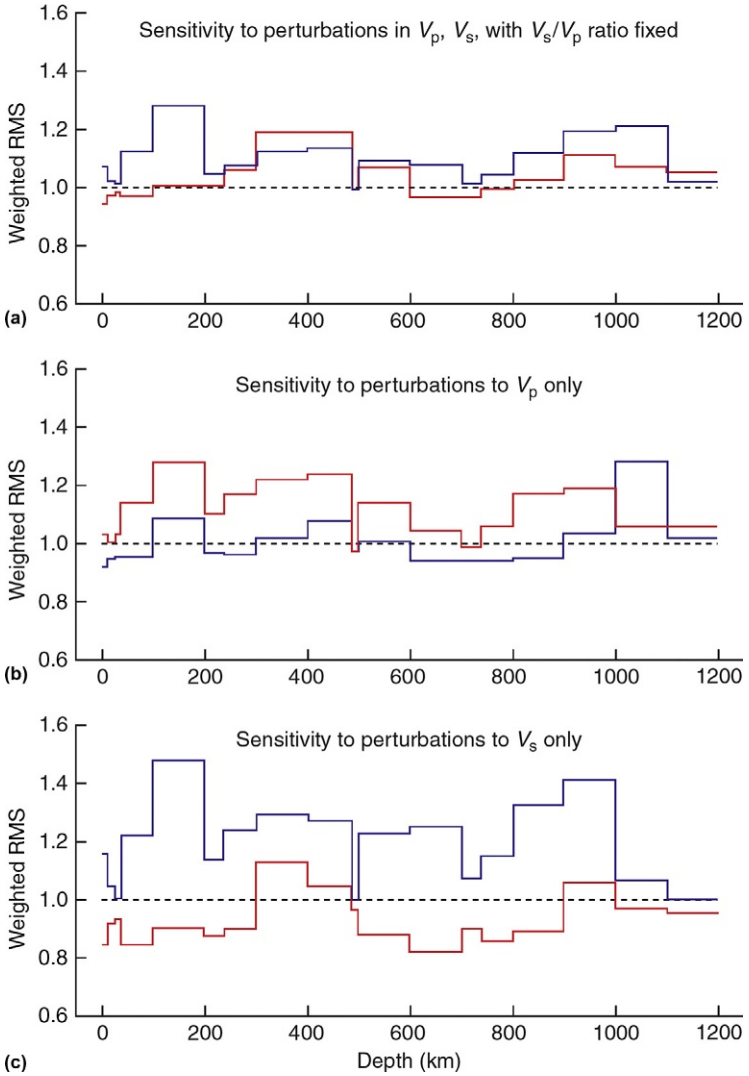


Figure 7 Weighted RMS misfit of the predicted travel times to the observed travel times (Lognonné et al., 2003) for velocity models in which GB06 is perturbed by 10% in a specified 100-km-thick interval. Where the published velocity model comprises layers of thickness <100 km, the published layer thickness is used. Travel-time residuals (observed minus predicted) are weighted by the uncertainty in the travel-time pick, and the total misfit is normalized such that the misfit to the unperturbed GB06 is 1 (dashed line). Red – GB06 increased by 10% as a function of depth (red); blue – GB06 decreased by 10%. (a) The v_p/v_s ratio is held constant, as specified in GB06, the P-wave velocity is perturbed by $\pm 10\%$ in a specified depth interval, and v_s is perturbed using the v_s ratio. RMS misfit is computed using both P and S arrival times. (b) Only v_p is perturbed by $\pm 10\%$, and the misfit to the P-wave arrival times computed. (c) Only v_s is perturbed by $\pm 10\%$, and the misfit to the S-wave arrival times computed.

and models with 10% increases or decreases in velocity provide poorer fits to the data. Sensitivity at depths of 500–800 km is more limited, however, and, in particular, higher velocities can be tolerated with little change in RMS misfit. Sensitivity in the region of deep moonquakes again increases, with lower velocities resulting in a poorer fit to the data.

In summary, the lunar seismic events recorded by the Apollo network have resulted in travel-time data sets (Lognonné et al., 2003; Nakamura, 1983) that enabled 1D seismic-velocity profiles to be inferred to depths of about 1200 km. Recent work suggests a revision of crustal thickness from previous estimates of ~60 km (Nakamura, 1983) (an average value, probably most representative of the thickness in the Apollo station 12 and 14 region) to <45 km (Khan et al., 2000). The lower thickness value could be around 38 km, as proposed by Khan and Mosegaard (2002), or it could be as low as 30 km (Gagnepain-Beyneix et al., 2006; Lognonné et al., 2003). While both models are compatible with the receiver-function arrival times (Vinnik et al., 2001), models of receiver-function amplitudes favor a thin crust with low seismic velocities. Upper mantle seismic-velocity estimates are generally in agreement. Sensitivity to seismic velocity in the midmantle (300–700 km) is limited and critically depends on the crustal structure. Thus, a range of models is permitted, including those with small-amplitude negative seismic-velocity discontinuities or gradients at ~300 km depth and those with sharp positive discontinuities at ~500 km depth (**Figures 6** and **7**). Seismic-velocity estimates in the deep-moonquake zone differ, with P-wave velocities as high as 10 km s^{-1} obtained as a result of some travel-time data inversions

(Khan and Mosegaard, 2002; Khan et al., 2000), and all models show velocities higher than those in the upper mantle.

In **Section 10.03.2.4,** we summarize the constraints that the seismic models place on lunar thermal and mineralogical structures. Much of the remaining disagreement among 1D velocity models likely results from errors in arrival-time determinations, lack of resolution, differences in inversion techniques, and averaging of 3D structure. Despite initial attempts to use 3D tomography to determine mantle (Zhao et al., 2008, 2012) and crustal thickness (Chenet et al., 2006), and even with more rigorous modeling techniques (e.g., Wang et al., 2013a,b), four stations provide very limited 3D tomography with nonlocated sources. Thus, the next generation of 3D models will depend on constraining the location of the largest meteorite impacts by identifying of the freshest crater with LRO (Lunar Reconnaissance Orbiter) data. The low resolution of the Apollo seismometers and their limited bandwidth may also lead to a misinterpretation of the P and S arrival times, if the first arrivals are too small to be detected, as illustrated by Nakamura (2005), who studied the deviation between five arrival-time data sets (Goins, 1977; Lammlein, 1977; Lognonné et al., 2003; Nakamura, 1983, 2005). Differences between these data sets are large – often 10 s or more – and illustrate the difficulties of phase-picking in narrow band records of scattered waves. KM02, GB06, and N83 also employ differing inversion procedures, especially with regards to the relocalization procedure performed after each iteration. In view of the combined effects of 3D structure and poor seismic sampling of the lunar interior due to the limited source–receiver geometries, new data provided by future networks with very-broad-band (VBB) seismometers are required. Therefore, several projects in Japan (e.g., SELENE2, Tanaka et al., 2013), the United States (e.g., Lunette, Neal et al., 2011), and Europe (e.g., FARSIDE, Mimoun et al., 2012) have been proposed for establishing a new seismic station(s) on the Moon.

10.03.2.3 Very Deep Interior

Many geophysical studies indicate that the Moon has a core (for a review see Hood and Zuber, 2000), which is evidenced by magnetic induction signatures (Hood et al., 1999) or remanent magnetism (Hood, 1995; reviewed in Cisowski et al., 1983; Fuller and Stanley, 1987). Geochemical analyses of mare basalt samples indicate a depletion of highly siderophile elements (e.g., Righter, 2002) relative to the depletion level expected from any lunar-core formation scenario (Canup and Asphaug, 2001). Impact simulations (see Cameron, 2000) suggest that a low fraction of iron from the proto-Earth and proto-Moon was put into orbit after the giant impact. These mass-fraction estimates are typically 1% or less, and they reach 3% in only a few extreme cases, as iron can be further added during late-stage accretion.

Until recently, the only methods for directly investigating the lunar core were magnetic sounding and geodesy. Magnetic sounding (Hood et al., 1999) is based on the induced magnetic dipole moment produced by the motion of the Moon through the Earth's geomagnetic tail. A core radius of 340 ± 90 km is inferred by this method, under the assumption that electric currents in the core can be approximated by a current 'sheet' localized on the core surface. The second approach, the measurement of the moment of inertia ratio (0.3932 ± 0.0002, Konopliv et al., 1998), indicates that the density toward the center of the Moon is higher than it is inside the lunar mantle. Moreover, analyses of lunar rotation (Bois et al., 1996; Williams et al., 2001) have shown that the rotation of the Moon is influenced by a dissipation source, which has been interpreted as the signature of a liquid core.

A lunar core was also suggested by interior structure models obtained from inversions of the density, moment of inertia, Love number (k_2), and even the induction signature, with or without the additional constraints provided by the seismic data. Bills and Rubincam (1995) used the mean density and the inertia factor only and they estimated a core radius of 400 and 600 km, respectively, for densities of 8000 and 6000 kg m^{-3}. Khan et al. (2004) used these constraints, along with the Love number, and performed a Monte Carlo inversion assuming a 5-shell model. The inversion inferred a core with a radius of about 350 km and a density of 7200 kg m^{-3}. As several tradeoffs exist between the size and density of these layers, the independent constraints from seismology can be added in order to limit the space of acceptable models. Interior structure inversions based on a priori seismic models were first performed by Bills and Ferrari (1977), using a preliminary seismic model, and later by Kuskov and Kronrod (1998) and Kuskov et al. (2002), using Nakamura's (1983) seismic model. Kuskov and colleagues proposed either a pure γ-Fe core with a density of 8100 kg m^{-3} and a radius of 350 km or a core with smaller densities and larger radius, including the largest troilite FeS core with a radius of 530 km and a density of 4700 kg m^{-3}. Khan et al. (2006) performed another study using seismic information, the inertia factor, and the mean density, and he predicted a core with a density of about 5500 kg m^{-3}.

The geometry of the lunar network, particularly, its lack of any antipodal stations mean that the system has recorded few, if any, ray paths propagating deep in the Moon (>1200 km depth) (**Figure 6**; see Nakamura et al., 1974b for an impact on the far side). Therefore, the body's core cannot be geometrically determined by direct waves (e.g., Knapmeyer, 2011). An alternative approach for investigating core structure involves exploring a planet's normal modes (e.g., Lognonné and Clévédé, 2002). A search for free oscillations in the Apollo data has been performed by a few authors, as low-angular-order normal modes are sensitive to core structure. After an unsuccessful attempt by Loudin and Alexander (1978), Khan and Mosegaard (2001) claimed detection of free oscillations from flat-mode LP Apollo signals generated by meteorite impacts. However, Lognonné (2005) and Gagnepain-Beyneix et al. (2006) have shown that the signal-to-noise ratio of these events was probably too small to result in detectable LP signatures. Nakamura (2005) has suggested the presence of about 30 possible deep moonquake source regions on the lunar farside: however, no events were detected within 40° of the antipode of the mean sub-Earth point, suggesting that this region is either aseismic or strongly attenuates or deflects seismic energy (Nakamura, 2005; Nakamura et al., 1982).

Two recent studies have independently reanalyzed Apollo seismograms using modern waveform methods to search for reflected and converted seismic energy from a lunar core, by using stacking methods. The success of these two analyses can

be understood through Figure 8(a), which shows the amplitudes of the individual deep-moonquake P and S waves, on which the typical amplitudes of core phases have been superimposed (for simple isotropic sources). This illustrates the relatively low amplitude of the ScS phases with respect to the instrument detection threshold and suggests the possibility of signal enhancement by stacking. These stacks form the basis for the search performed by the two separate studies of Weber et al. (2011) and Garcia et al. (2011).

Weber et al. (2011) used polarization filtering (similar to the double-beam stacking method in terrestrial array seismology) to attempt to identify reflected core phases (PcP, ScS, ScP, and PcS) from three deep lunar interfaces: the top of a partial-melt layer at the base of the mantle, the interface between an outer fluid core and the lower mantle partial-melt layer, and the interface between an inner solid and outer fluid core. P- and S-wave velocities in the layers, as well as the radius of the interfaces, were determined. The resulting model is one in which the top of the partial-melt layer lies at a radius of 480 ± 15 km, and the tops of the outer and inner core are at 330 ± 20 and 240 ± 10 km, respectively. The inferred solid- and liquid-core radii suggest a core that is 60% liquid by volume, and these measurements constrain the concentration of light elements in the outer core at less than 6 wt%. Garcia et al. (2011) constructed a 1D reference Moon model incorporating both seismological and geodetic (density, moment of inertia, and Love number (k_2)) constraints. First, radial variations in P- and S-wave velocities and density that match the seismic and geodetic data were inverted for different values of core radius. Then, using waveform stacking and a polarization filtering technique, but also taking into account a correction for the gain of the horizontal sensors, a best-fit core radius was determined. Garcia et al. (2011) found a best-fit core radius of 380 ± 40 km, larger than the radius determined by Weber et al. (2011), thus permitting somewhat higher concentrations of light elements (up to 10 wt%) and a best-fit mean core density of 5200 ± 1000 kg m^{-3}, which is significantly different from the average density of the inner and outer core of 6215 kg m^{-3} found by Weber et al. (2011).

These two seismic analyses confirm the existence of the core, and both support a fluid outer core and a solid inner core. However, the uncertainties in core radius remain large, with estimates ranging from 300 to 400 km, and, in fact, most of the deep geophysical properties of the Moon are still weakly constrained. Figure 8(b) shows the typical range of several deep lunar parameters, such as mid- and lower-mantle density, lower-mantle shear-wave velocity, core radius, and core density. The inverse problem remains underdetermined (the data are k_2 Love number, density, moment of inertia factor, and the ScS travel time). The two seismic models of Weber et al. (2011) and Garcia et al. (2011) mainly differ in their treatment of the lower mantle structure. This structure is proposed to be a low-velocity, partially melted zone by Weber et al. (2011), in contrast to the model of Garcia et al. (2011), in which this zone has velocities close to those in the midmantle. These data and models all suggest a core that comprises 0.75–1.75% of the lunar mass with a mean density less than 6215 kg m^{-3}, consistent with the presence of some light element(s). This is also consistent with estimates of the temperature at the core–mantle boundary, which are compatible with a liquid core only if the latter contains light elements (Gagnepain-Beyneix et al., 2006; Khan et al., 2006; Lognonné et al., 2003). A core with little or no light elements, corresponding to the high densities found by Khan et al. (2004), will likely be solid at those temperatures and can be excluded.

More precise deep-interior-structure estimation will depend on new geophysical data and (even independent of these) on better estimation of the thermal state of the lunar lower mantle. Such constraints might possibly be obtained from deep-moonquake dynamics, as they provide another important constraint on deep lunar structure. Density and elastic moduli from the seismic models can indeed be used to explore tidal stresses as a function of depth (Figure 9) and/or time (Bulow et al., 2006). Understanding tidal stresses as a function of time and position is critical to understanding how and why deep moonquakes occur, because the seismic data distribution and quality prohibit the inference of focal mechanisms for these events.

10.03.2.4 Mineralogical and Thermal Interpretation of Lunar Seismic Models

Considerable effort has been expended in interpreting the depth dependence of lunar seismic velocity in terms of mineralogical and thermal structure. An excellent review of this topic has been given recently by Wieczorek et al. (2006), and so we focus here on only the major results and outstanding issues from these studies. Two general approaches have been used: investigations of a limited suite of compositional/mantle mineral assemblages (e.g., Gagnepain-Beyneix et al., 2006; Hood and Jones, 1987; Lognonné et al., 2003; Mueller et al., 1988), and a more complete thermodynamic treatment (series of papers by Kuskov and colleagues, e.g., Johnson et al., 2005; Khan et al., 2006, 2007; Kuskov, 1995, 1997; Kuskov and Fabrichanaya, 1994; Kuskov and Kronrod, 1998, 2009; Kuskov et al., 2002). In the latter approach, a bulk composition is specified as a function of depth, and the equilibrium mineralogical assemblage is calculated for a given selenotherm (lunar temperature profile), from which the elastic moduli and resulting seismic velocities can be predicted. An example is the model of Kuskov et al. (2002) in Figure 5. Seismic velocity models predicted from mineralogy can then be compared with those inferred from travel-time data, and the mass and moment of inertia can be compared with geodetic constraints. Alternatively, travel times can be calculated using the predicted seismic-velocity models and compared directly with the observations (Johnson et al., 2005; Khan et al., 2006, 2007). Of particular interest for the Moon have been (i) the bounds that crustal thickness (through the concentration of heat-producing elements in the crust) and mantle seismic velocities place on temperature, (ii) whether mantle seismic velocities can discriminate among broad classes of interior differentiation models, and (iii) whether phase changes are identifiable in the seismic velocity structure.

Studies to date (Hood and Jones, 1987; Khan et al., 2006; Mueller et al., 1988) show that no single compositional model predicts seismic velocities that match both the upper and lower mantle 1D profiles, even when temperature and pressure effects are accounted for. In general, models that assume

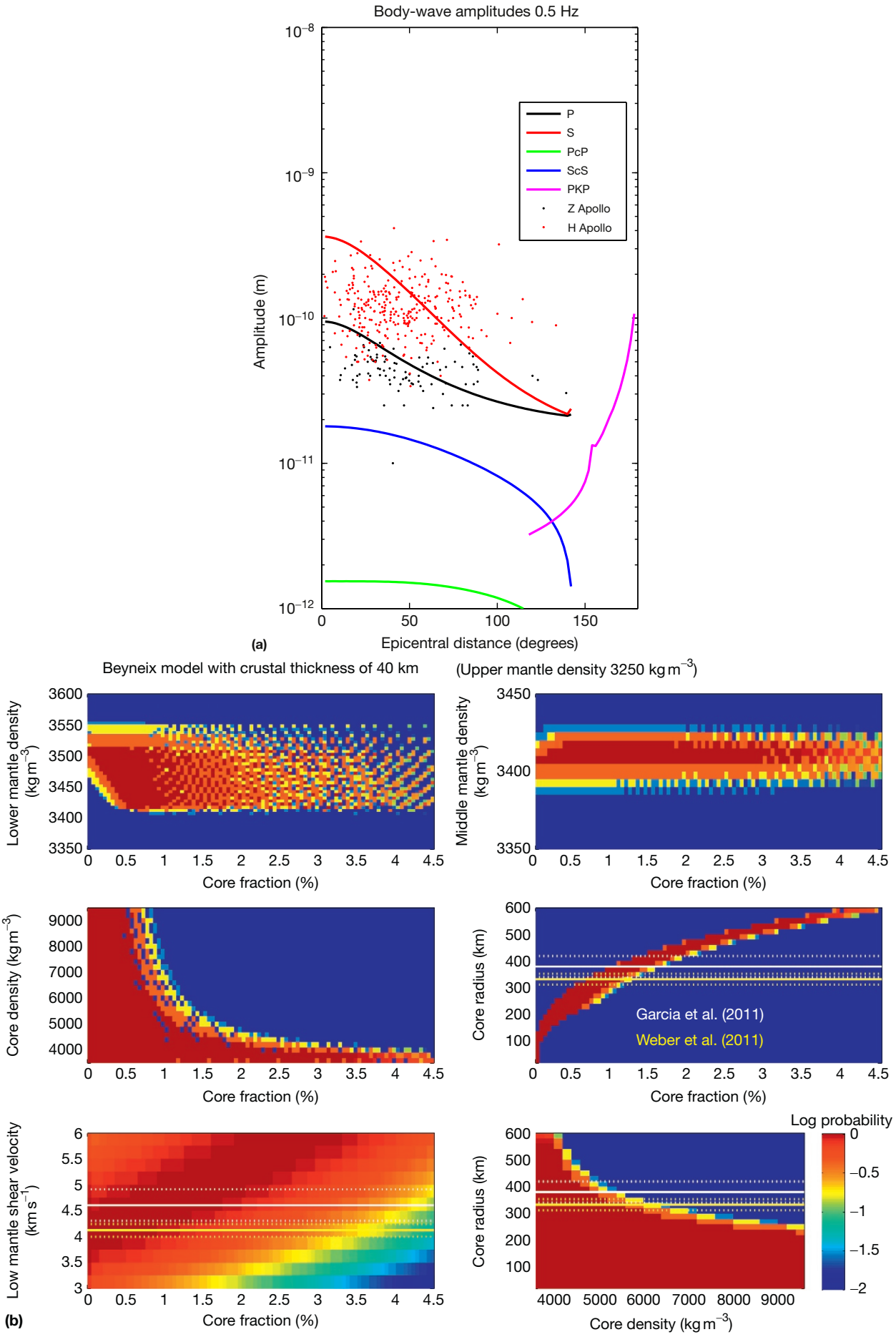


Figure 8 (Continued)

differentiation of the upper mantle and that are more aluminous in bulk composition than the terrestrial mantle are favored. As most mineralogical studies compared their predicted seismic velocities with the Nakamura (1983) model, a focal point was the presence or absence in the predictions of a velocity discontinuity at 500 km. Both phase changes (Hood and Jones, 1987; Mueller et al., 1988) and compositional boundaries (Kuskov, 1997; Mueller et al., 1988) at this depth were inferred. Compositional boundaries were of interest as they may reflect the initial compositional zoning of the Moon, its basis in an early lunar magma ocean, its resulting crystallization sequence, and the maximum melting depth of the mare basalt region (see discussion in Wieczorek et al., 2006). The discussion in **Section 10.03.2.2** indicates that the arrival times

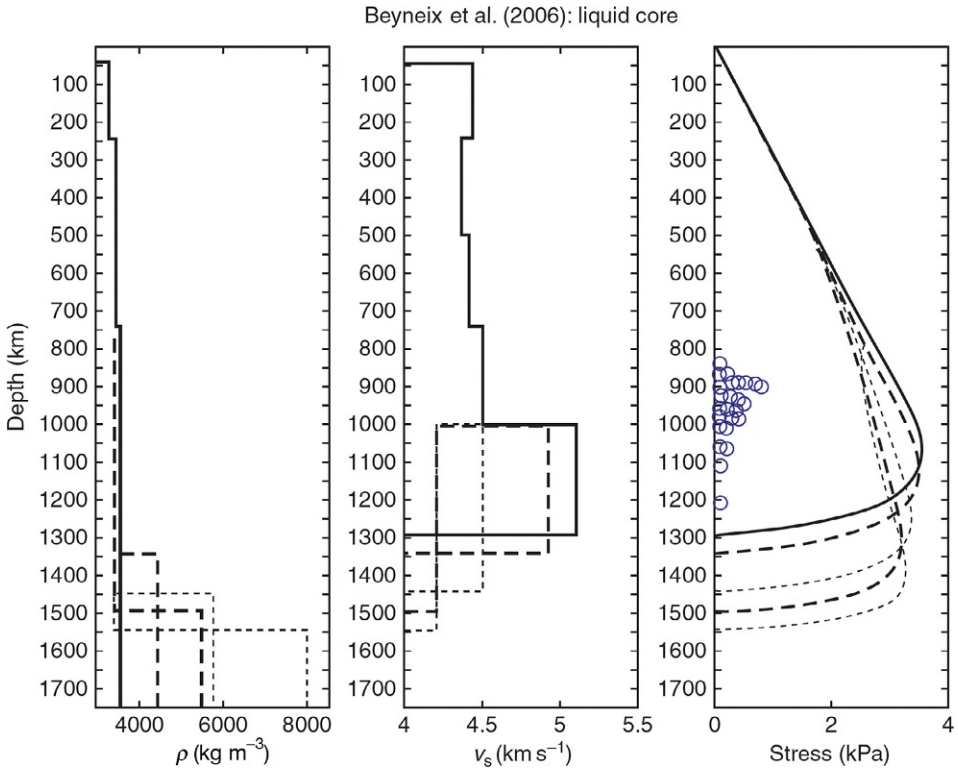


Figure 9 More specific illustration of the trade-off between lower-mantle rigidity and core radius. All seismic and density models shown match the Apollo seismic travel times, the mean density, moment inertia, and k_2 within the data error bars. All models have S-wave velocity values calculated by Gagnepain-Beyneix et al. (2006) for the mantle and crust, and only the shear velocity in the very deep mantle is modified. Shear velocity in the core is zero, as only models with liquid cores are shown. The various lines are each associated with a given core size and minimize the variance. From left to right, the figures represent the density, the shear-wave velocity, and the tidal stresses. Models with the largest cores (400 km or more) correspond to an ilmenite core with densities lower than 5000 kg m^{-3}. These models have correspondingly high shear velocity in the lower mantle. Models with a core radius of 350 km correspond to an FeS core, with densities in the range of 5000–6000 kg m^{-3}. Smaller cores (~200 km) of larger densities are also compatible with the data, if associated with a low-velocity zone in the lower mantle, in order to match the low k_2 value. On the far right, the maximum horizontal tidal stress is shown with respect to the depth and defined as $(T_{\theta\theta}+T_{\phi\phi})/2$, where T is the tidal stress tensor, at the latitude and longitude of the deep Moonquake A1 (as found by Gagnepain-Beyneix et al. (2006), i.e., -15.27° S, -34.04° E). See Minshull and Goulty (1988) for more details on stress computations. Note that only models with a core radius of 350 km or more produce maximum tidal stresses in the vicinity of the deep moonquakes.

Figure 8 (a) Typical amplitudes of the P and S body waves of deep moonquakes detected by Apollo, as a function of epicentral distance. Z Apollo are the amplitudes recorded for P on the vertical axis of Apollo, while H Apollo are those for S on the horizontal axis. The amplitudes are taken from the Nakamura et al. (2008) catalog, but converted to displacement using conversion factors between mm and peak-to-zero displacement obtained through the comparison of the catalog amplitudes with amplitudes recorded by the A1 deep-moonquake seismograms after instrument correction. As an indication, typical relative amplitudes of P, S, and core phases (ScS, PcP, and PKP) are plotted for the interior model of Garcia et al. (2011), illustrating that ScS amplitudes, although too small to be individually detected in the Apollo data, might be detected through stacking for the largest events. PcP phases have amplitudes that are too small to be identified with stacking, however, and they will remain challenging, even for the next generation of lunar seismometers. (b) Exploration of the model space for acceptable models for density, inertia factor, and Love number k_2, using the seismic models of Gagnepain-Beyneix et al. (2006), compared to the core estimations of Garcia et al. (2011) and Weber et al. (2011), which are represented by white and yellow lines, respectively. The middle mantle is defined as occurring between 1500 and 1000 km radius, while the lower mantle occurs between 1000 km and the core radius. The color scale represents the decimal logarithm of exp(−var), proportional to the probability, where the variance is between the computed and observed densities, moment of inertia, and k_2. For the variance definition, values, and errors, see Khan et al. (2004). Acceptable models are dark red and red. The model space is sampled in order to identify the range of acceptable solutions. In the middle–lower mantle S-wave velocities are equal to 4.5 km s^{-1} in the models of Gagnepain-Beyneix et al. (2006), while the mean velocities of Garcia et al. (2011) and Weber et al. (2011) are 4.6 and 4.125 km s^{-1}, respectively.

show limited sensitivity to velocity structure at depths of 300–500 km, and even up to 800 km, and the section cautions against overinterpretation of mantle structure in this region. Overall, however, the increase in seismic velocity from the uppermost mantle (0–300 km) to the region of deep moonquakes (700–1200 km) is consistent with a change in bulk composition from a dominantly orthopyroxene upper mantle (with lesser amounts of olivine, clinopyroxene, plagioclase, and garnet) to a dominantly olivine lower mantle (with smaller amounts of garnet and clinopyroxene) (Wieczorek et al., 2006). This kind of compositional change is associated with an increase in magnesium number with depth. The studies by Lognonné et al. (2003), Gagnepain-Beyneix et al. (2006), and Khan et al. (2006) demonstrate the broad range of acceptable models, given current seismic constraints.

Crustal thickness plays an important role in determining the thermal profile of the lunar interior due to the preferential sequestration of heat-producing elements in the crust. Again, a large range of thermal profiles and resulting mineralogical assemblages are compatible with the seismic data (Khan et al., 2006), but recent studies indicate a preference for 'cold' thermal profiles (Gagnepain-Beyneix et al., 2006; Lognonné et al., 2003), with a temperature of 1073 K at 340-km deep (equivalent to the base of the terrestrial elastic lithosphere) and a temperature of 1473 K at 740-km deep (base of terrestrial thermal lithosphere). Such depths are comparable to the depths found in thermal evolution models of the Moon (e.g., Spohn et al., 2001a). In these models, bulk lunar abundances of the heat-producing elements uranium and thorium are similar to terrestrial values, with an enriched crust and a depleted mantle. For example, mantle abundances of about 8.2 ppb of U and 30 ppb of Th are suggested by Gagnepain-Beyneix et al. (2006) in order to fit the temperature dependence of seismic velocity. These values of U and Th abundances are close to those proposed by Waenke et al. (1977) and Taylor (1982).

Taken together, mineralogical and thermal studies of the lunar interior, using seismic data, show that a broad range of interior models are compatible with that data. Yet, the existing models still have limitations, including a lack of detail regarding thermodynamic parameterization and the absence of titanium (and sometimes sodium) from calculations. Furthermore, the likelihood of 3D structure in the Moon is great, as suggested by petrological models (e.g., Elkins-Tanton et al., 2011), the possible spatial association of deep moonquakes with the lunar mare (Qin et al., 2012; Watters and Johnson, 2010), and recent GRAIL results that clearly show spatial variations in crust density and porosity (Wieczorek et al., 2013). Thus, the broad range of acceptable mineralogical and thermal models may simply reflect 3D variations in the seismic data that have been mapped into 1D structure (necessarily, given the data set).

10.03.3 Seismic Activity of the Moon and Terrestrial Planets

10.03.3.1 Internal Seismic Activity

So far, no conclusive indication of present-day plate tectonic activity has been observed on a planet other than Earth. Moreover, only the Moon's seismic activity is constrained by seismic data, and the seismic activity of Mars and Venus can only be estimated from models based on faults density or theoretical thermoelastic cooling. During the Apollo seismic network's 7-year operating life, about 13 050 seismic signals were detected on the LP instruments and cataloged, and many more events recorded by the SP instruments remain uncataloged. **Figure 10** shows the annual detection statistics on the horizontal component of the Apollo 14 station, based on Nakamura's catalog. We focus first on the internal activity of the Moon, and then we address the internal activity of other planets, followed by a discussion of impact seismology.

Assuming that the largest deep moonquakes have a seismic moment of 5×10^{13} Nm and associated stress drops of about 10 kPa (Goins et al., 1981a), quakes with seismic moments 30 times smaller were detected. These smallest reported moonquakes correspond to terrestrial events with body-wave magnitudes as low as 1.6 and are detectable because of the low seismic noise level on the Moon (**Figure 1**). Note, however, that the definitions of the body- or surface-wave magnitudes are generally based on Earth conditions. Therefore, unless specified differently, we will use moment magnitudes directly related to the seismic moment by $M_w = \frac{2}{3}(\log_{10}(M_0) - 9.1)$, where M_w is the moment magnitude and M_0 the seismic moment in Nm (relation (4) of Hanks and Kanamori, 1979 after unit corrections).

Deep moonquakes originate from regions that appear to undergo repeated failure, giving rise to sets of moonquakes with similar waveforms and periodic occurrence times (Lammlein et al., 1974; Nakamura, 1978; Toksöz et al., 1977 and more recently, Frohlich and Nakamura, 2009; Weber et al., 2009). The number of known source regions for deep moonquakes is currently estimated at ~250 (Nakamura, 2003, 2005), and source depths range from 700 to 1200 km. **Figure 11(a)** shows the variation in the number of moonquakes recorded per week for the duration of the Apollo seismic experiment. The installation dates of the stations are provided in **Table 1**. The activity at all known moonquake nests (reported in Nakamura et al., 2008) is shown in red, and the activity recorded at nine clusters that dominate the catalog is shown in blue. **Figure 11(a)** illustrates that the activity at these nine well-studied clusters closely resembles the behavior of the larger deep-moonquake population. Individual peaks in the time series occur at approximately 2- and 4-week intervals. Also apparent is the modulation of moonquake activity during the 206-day period, and possibly during a period close to the duration of the experiment (6 years). **Figure 11(b)** shows the power spectrum of moonquake activity. The number of moonquakes per 24-h day was calculated, and the power spectrum was computed using the minimum-bias multitaper method of Riedel and Sidorenko (1995). A multitaper method was chosen because tidal effects with periods of 206 days and 6 years are suggested by the time series in **Figure 11(a)**. The 6-year period, in particular, requires the full length of the time series (8 years) to be analyzed. Various numbers of tapers were explored. **Figure 11(b)** shows the power spectrum computed using six tapers. This number allows frequency resolution at the expense of noisy spectral estimates, as is evident at high frequencies. Spectral peaks are seen at 13–14 and 27–28 days, with some structure seen in the monthly peak. A broad peak in the power spectrum is seen close to the previously noted 206-day period, and the increase

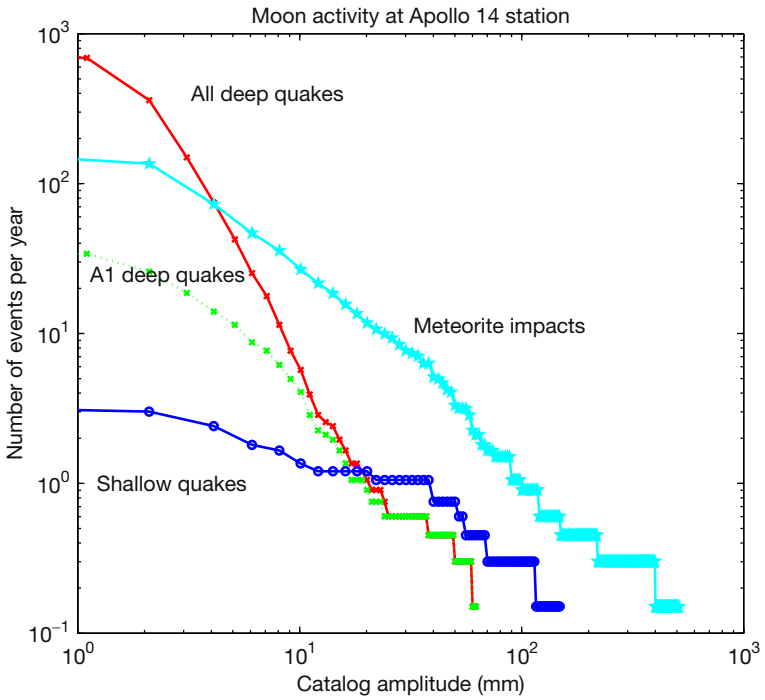


Figure 10 Number of events per year detected by the Apollo 14 station versus amplitude. The amplitudes are found in the Nakamura et al. (2008) catalog. The conversion of amplitudes in physical units depends on the instrument function response and on the dominant frequency. For a pure sinusoidal signal, the amplitude x (in mm) is about half of the amplitude of the time derivative of the signal y (in DU/s) (Nakamura, personal communication). At 1 Hz, which is the dominant frequency of the deep moonquakes recorded at Apollo 14, the resolution of the peaked mode is about 3.5×10^{-10} m, and a distance of 1 mm therefore corresponds to about 10^{-10} m in ground displacement. For high-frequency teleseismic and impacts, the scaling depends on the distance. Note the very high number of events detected per year in all cases, despite their very low amplitudes (see **Figure 1** for seismometer sensitivity).

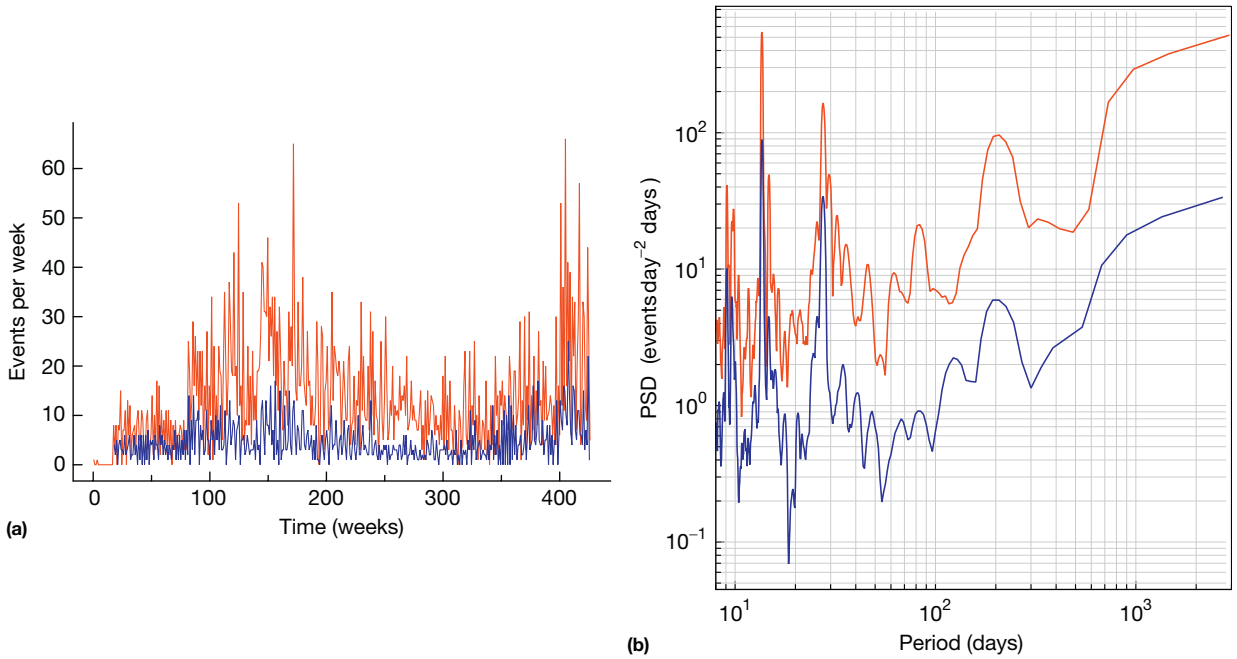


Figure 11 Deep-moonquake activity. (a) Number of moonquakes per week versus time in weeks, starting from the beginning of the seismic experiment. Moonquake activity, which is based on a recent update of the deep moonquake catalog (Nakamura et al., 2008), is shown in red. For comparison, the moonquake activity at nine of the most active deep clusters (A1, A6, A8, A9, A10, A14, A18, A20, and A33) is shown in blue. This moonquake activity includes new events recently found by Bulow et al. (2005, 2006) and not included in the lunar catalog (Nakamura et al., 2008). (b) Power spectral density (PSD) of the number of moonquakes per day versus period in days. Peaks at ~14, 28, and 206 days can be seen, and the power at the longest periods likely results from a 6-year modulation of moonquake occurrence times. Red and blue curves are as in (a).

in power at the longest periods likely reflects the 6-year modulation of moonquake occurrence times (Lammlein, 1977; Lammlein et al., 1974). Energy from this period is spread over the five lowest frequency spectral estimates, which occur at periods of 2916, 1458, 972, 729, and 583 days. Increased smoothing using more tapers or the fully adaptive multitaper approach results in spectra with broader peaks at the monthly and fortnightly periods and the loss of any resolution at around 200 days. Thus, this spectral approach, which is typically used in investigating tidal periodicities in terrestrial quakes (e.g., Tolstoy et al., 2002), provides only crude insight into the lunar data. In contrast, the spike train approach of Bulow et al. (2007) allows resolution of the distinct contributing monthly periods.

In addition to the deep quakes, 28 other events were detected at much shallower depths. These events may be analogous to terrestrial earthquakes. They show no obvious correlation with tidal activity or with surface features on the Moon, although no events in the southeast quadrant of the Moon's nearside were detected. (The latter observation is highly uncertain due to the paucity of observed shallow events.) All focal depths for the shallower events are <200 km. The largest shallow events have a seismic moment of about 3×10^{14} Nm for Goins et al. (1981a) or 1.6×10^{15} Nm for Oberst (1987), and the equivalent moment magnitudes range from 3.6 to 4.1. Their depths have been estimated to be <200 km, and they are interpreted as resulting from the release of tectonic stresses (Nakamura et al., 1982). Even the SP seismometers recorded flat displacement spectra, indicating a corner frequency higher than 10 Hz in most the cases (Goins et al., 1981a). This implies very high stress drops, ranging from less than 40 MPa for Goins et al. (1981a) to 210 MPa for Oberst (1987), perhaps due to the cold, volatile-poor (and hence rigid) lunar lithosphere. Interestingly, these events resemble Earth's intraplate earthquakes, which also exhibit stress drops (for a given seismic moment) much higher than the drops associated with plate boundary-related earthquakes (Scholz, 1990). (Intraplate earthquakes have been used as a reference for all estimates of seismic activity on the telluric planets).

We now examine how quakes can be released by the thermoelastic cooling of the lithosphere. We apply the theory to the Moon in order to possibly explain the shallow lunar events, and to Mars in order to predict possible background seismic activity. The cumulative seismic moment, over a time Δt, is given by

$$M_{\mathrm{cum}} = \eta \dot{\varepsilon} V \mu \Delta t \qquad [1]$$

where V is the seismogenic volume, η is the seismic efficiency, and μ is the mean shear modulus of the seismogenic layer (Bratt et al., 1985; Phillips, 1991). Following Phillips (1991), the strain rate $\dot{\varepsilon}$ is given as

$$\dot{\varepsilon} = \frac{1}{H\left(a-\frac{H}{2}\right)^2}\int_{a-H}^{a} \alpha \dot{T} r^2 \mathrm{d}r \qquad [2]$$

where H is the thickness of the seismogenic layer, α is the thermal expansion coefficient, $\dot{T}$ is the cooling rate, and a is the planetary radius. If we assume that the thermal gradient in the seismogenic layer is linear and given by $T=T_0+((a-r)/H)\Delta T$, and that H thickens with time at a rate, $\dot{H}$ (both $\dot{T}$ and $\dot{H}$ are counted positively), we then have

$$M_{\mathrm{cum}} = 2\pi\alpha\eta\mu a^2 H \Delta T\left(\frac{\Delta\dot{T}}{\Delta T}+\frac{\dot{H}}{H}\right)\beta\Delta t \qquad [3]$$

where ΔT is the contrast of temperature between the bottom of the seismogenic layer and the surface, and where

$$\beta = \left(1-\frac{H}{a}+\frac{1}{3}\frac{H^2}{a^2}\right)\frac{1-\frac{4}{3}\frac{H}{a}+\frac{1}{2}\frac{H^2}{a^2}}{1-\frac{H}{a}+\frac{1}{4}\frac{H^2}{a^2}} \qquad [4]$$

is a geometrical factor correcting the thin-shell approximation. This factor is greater than 0.95 for Mars, the Earth, and Venus, and it is 0.72 for the Moon.

For the Moon, we have assumed $H=400$ km, which is an estimate of the depth of the 1073 K isotherm (Spohn et al., 2001a). The isotherm corresponding to the base of the lunar seismogenic layer is unknown. Abercrombie and Ekström (2001) take 873 K as the limit of brittle failure in oceanic crust, and this temperature is reached at depths of about 250–300 km on the Moon, using the thermal models of Spohn et al. (2001a) and Gagnepain-Beyneix et al. (2006), respectively. None of the shallow moonquakes occurred at depths >200 km, however. On the other hand, the lunar crust might be rich in anorthite, and, in that case, the pressure and temperature at that depth might encourage the transition of feldspar from brittle to plastic, as happens at depths of 20–30 km in the Earth's continental crust (Scholz, 1990). Deep moonquakes occur at much greater depths, up to 1100–1200 km, where the temperature seems high (1600–1700 K) according to all published models. This relationship between deep-earthquake frequency and higher temperatures remains a paradox, either related to our models of the lunar interior, lunar mineralogy, or our current understanding of fault rupture under these conditions. Yet, as noted above, the energy released by deep moonquakes is limited compared to the energy released by shallow quakes.

For Mars, Phillips (1991) has assumed a constant cooling rate of $\dot{T}$ equal to 1.1×10^{-7} K year^{-1}, and using a thermal evolution model, Knapmeyer et al. (2006) have set the $\dot{T}$ value at 0.5×10^{-7} K year^{-1}. Taking the 873 K lower bound on the isotherm at the base of the seismogenic layer, and the thermal evolution models of Spohn et al. (2001b), we estimate $\dot{H}$ to be in the range 3.15×10^{-8} to 4×10^{-8} km year^{-1}. (For comparison, this is 10–20 times smaller than the rate of thickening of the 100 Ma oceanic lithosphere.) Using $\alpha=2\times10^{-5}$ K^{-1}, $\mu=66$ GPa (obtained for $v_s=4.5$ km s^{-1} and $\rho=3300$ kg m^{-3}) and 3.5×10^{-8} km year^{-1} for $\dot{H}$, $M_{\mathrm{cum/year}} = a\mu\Delta t 2\pi a^2 \dot{H}\Delta T \approx 5.5\times10^{17}$ Nm year^{-1}, a value to be compared with the annual release of the shallow quakes (7.3×10^{14} Nm year^{-1}). Such a rate is similar to the one obtained if constant $\dot{T}=0.5\times10^{-7}$ Kyear^{-1} is assumed for a 200-km seismogenic layer. The ratio between observations and the model is about 1.3×10^{-3}, 20 times smaller than the lowest seismic efficiency parameters η reported for Earth (e.g., Ward, 1998). Thus, the shallow seismic energy release of the Moon is an outstanding puzzle: calculations of the type above, modeled on the approach of Phillips (1991), are either inappropriate or require an extreme choice of parameters relative to our current understanding. Lunar seismic efficiency may be lower than on Earth, possibly due to a cold, volatile-free seismogenic layer

able to support large initial stress without rupture. On the other hand, if these calculations are correct, we cannot exclude shallow moonquakes of much greater magnitude, but with period of recurrence longer than the 7 years of operation of the Lunar Network.

The seismic activity of Mars is unknown and can be estimated only through a comparative approach involving the Moon, as shown in **Figure 12**. Researchers have proposed only one possible seismic event on Mars during the 19-month operation of the Viking Lander 2 seismometer (Anderson et al., 1977a,b). During most of the experiment, the seismometer signal was correlated with the wind-related vibrations and lift of the lander. (See Lognonné and Mosser (1993) for a detailed explanation, and Nakamura and Anderson (1979) for an application of the seismometer as a wind sensor.) As the proposed event was detected at a time when no meteorological wind data were recorded, this single detection is far from conclusive! Moreover, Goins and Lazarewicz (1979) have shown that, due to mantle attenuation on Mars, the 4-Hz frequency of the seismometer was not optimized for the detection of remote events, and martian seismic activity comparable to or lower than terrestrial intraplate activity was still compatible with this lack of detection.

Seismic activity lower than the Earth's intraplate activity is indicated by all available theoretical predictions. Golombek et al. (1992) use surface-fault observations, and Phillips (1991) uses a lithospheric thermoelastic cooling model to propose seismic activity on Mars 100 times greater than the shallow moonquake activity detected by the Apollo seismometers. The greater activity results from the effect of surface area (a factor of about 4), a seismogenic efficiency closer to Earth values (a factor of at least 20), and a possibly greater cooling rate (a factor of about 2). However, the amplitudes of the largest Mars seismic signals are still expected to be about four orders of magnitude lower than those of the largest earthquakes at LPs, that is, at frequencies below the source cutoff. **Figure 13** shows that martian seismicity might result in about 50 quakes with a seismic moment of 10^{15} Nm year^{-1}, with an increase–decrease of the quake frequency by a factor of 5 for a decrease–increase of the seismic moment by a factor of 10. This represents a cumulative seismic energy release of 10^{18}–10^{19} Nm year^{-1} (purple curve, **Figure 13**). These early estimates have been updated by Knapmeyer et al. (2006), with seismic activity ranging from a high seismic moment budget distributed over many events, to a low seismic moment budget distributed over a few events (**Figure 13**).

Although not modeled here, other sources of seismicity are possible. Volcanic activity, in the form of volcanic tremors or magma chamber retreat and associated faulting (e.g., as observed on Alba Patera by Cailleau et al., 2003), could lead to quakes, but although surface evidence for activity in the last 10 My has been proposed (Neukum et al., 2004), no evidence for present-day activity exists. The cooling of the most recently active volcanoes along with stresses associated with known gravity anomalies (e.g., at Argyre or Isidis (Zuber et al., 2000)) can supplement thermoelastic stresses. Landslides can also be associated with locally increased seismic activity. The distribution of potential tectonic quakes was studied by Knapmeyer et al. (2006) using the MOLA (Mars Orbiter Laser Altimeter) altimetry data. A total of about 7000 faults with a cumulative length of 600 000 km were found. Half of these faults appeared to be thrust faults, and the other half appeared to be normal, with no obvious correlation between fault density and ages. **Figure 14** shows the results of a Monte Carlo simulation of seismic activity on these faults, indicating

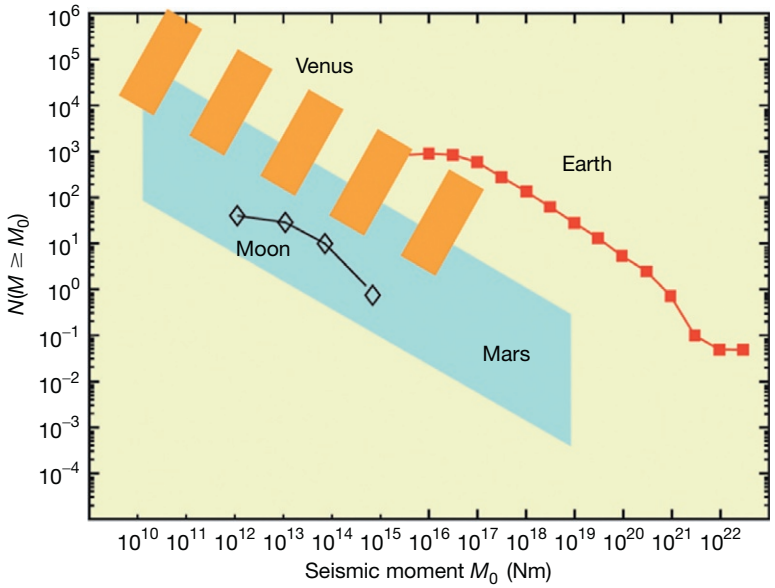


Figure 12 Seismic activity of the telluric bodies. The figure shows the annual number of events that are larger than a given seismic moment. For the Moon, the estimate is obtained from Oberst (1987) and only refers to shallow moonquakes. For Mars and Venus, the activity is estimated from various published models, based on the thermoelastic cooling of the lithosphere. The terrestrial activity is the mean activity for the period 1984–2004. The figure was adapted from results presented in Knapmeyer M, Oberts J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars' seismicity. *Journal of Geophysical Research* 111: E11006.

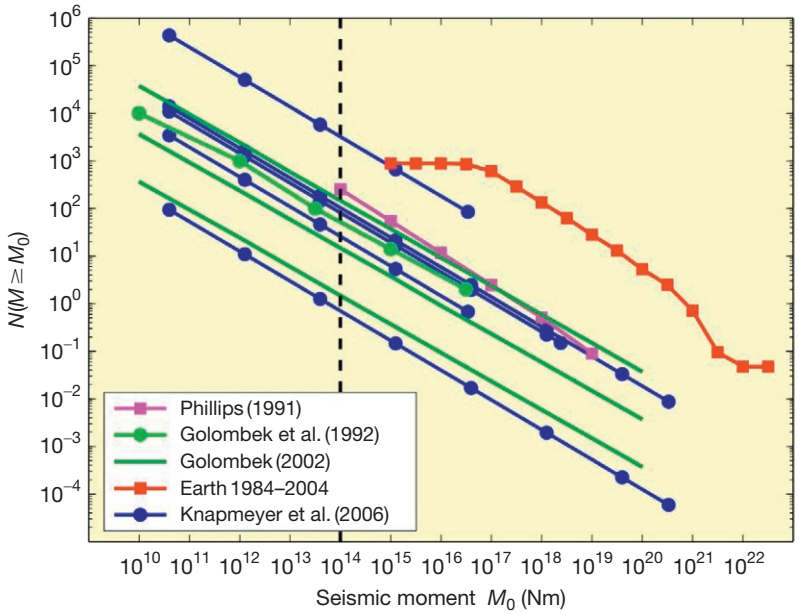


Figure 13 Comparison of models for annual martian seismic activity with the terrestrial mean annual activity (Knapmeyer, personal communication). Most models predict a total of 10–100 quakes per year with moments larger than 10^{15} Nm. The dashed line represents quakes of moment 10^{14}, for which S and P body waves have typical amplitudes of about 10^{-9} ms^{-2} in the 0.1–1 Hz bandwidth, with peak-to-peak distances of 50° and 70°, respectively (see **Figure 33(b)**). These waves may therefore be at the limit of detection for estimated martian noise levels (Lognonné et al., 2000). The two green lines shown for Golombek et al. (1992) and the blue line for Knapmeyer et al. (2006) refer to different hypotheses regarding the seismicity.

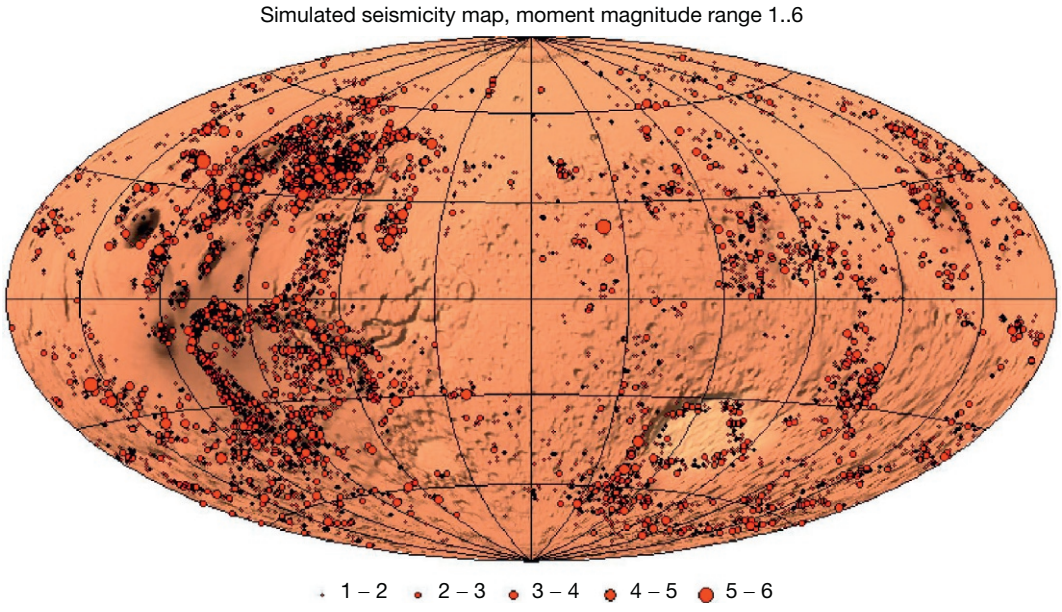


Figure 14 Seismicity map of Mars from Knapmeyer et al. (2006), determined using a Monte Carlo simulation of seismic activity releasing a cumulative moment of 10^{18}–10^{19} Nm per year (from Golombek et al., 1992), and with a probability distribution of epicenter locations determined by the density of surface faulting. The map shows the distribution of marsquakes that might be expected during one Earth year, with an annual seismic budget between the upper and lower bounds (Knapmeyer, personal communication).

that activity in the Tharsis region is very likely (Knapmeyer et al., 2006).

We briefly mention seismicity estimates for two other planets that might be targeted for seismic missions in the coming decades: Venus and the Jovian satellite Europa. Estimation of the seismic potential activity of Venus, performed as part of the NASA Venus Internal Structure Mission study (Stofan et al., 1993), suggested a seismically active planet. Assuming a seismogenic layer of 30 km, more than 100 quakes of $M_w > 5$ could be released by intraplate activity with a strain rate of 10^{-19} s^{-1} (Grimm and Hess, 1997). Quakes with $M_w > 6$ might be five times less frequent. A suggested rise in surface

temperature on Venus, over the period for which we have a geological record, generates compressive thermoelastic stresses in the crust (Dragoni and Piombo, 2003; Solomon et al., 1999). Such stresses might either generate reverse faults or act on preexisting reverse faults, and, by analogy with Earth, these stress might lead to quakes of maximum moment magnitude 6.5. A complicating issue on Venus is that of rheology; the absence of volatiles means that crustal rocks are much stronger than their terrestrial counterparts (Mackwell et al., 1998). Finally, as we note in our concluding remarks, there are severe technical challenges facing surface-based seismological experiments on Venus.

The seismic activity of the Jovian satellites is also unknown, although Io could be the most active of all known satellites in the Solar System (Kuskov and Kronrod, 2000) and a perfect target for a seismological mission. The possible activity of seismic activity on Europa has been reviewed by Panning et al. (2006), following studies of Kovach and Chyba (2001) and Lee et al. (2003) on the possibility of seismic experiments on this satellite of Jupiter. Possible seismic sources on Europa include the opening of tensile cracks due to the Jovian tides, as tide-induced cracks might lead to quakes ranging from 2 to 4 in moment magnitude. Nimmo and Schenk (2006) have also identified regions of normal faults on Europa by analyzing Galileo data, indicating potential quakes up to $M_w = 5.3$ magnitude. The frequency of such proposed events remains unknown.

10.03.3.2 External Seismic activity: Artificial and Natural Impacts

Impacts constitute about one-fifth of the seismic events detected by the Apollo network (1742 of the 9315 identified and classified events), and the strongest impact events have amplitudes comparable to the largest moonquakes. As discussed in **Section 10.03.2**, these surface seismic sources permit more detailed studies of crustal seismic structure, including lateral variations in thickness (Chenet et al., 2006).

The ability of meteorite impacts to generate seismic waves has been widely discussed regarding impacts on the Moon (Laster and Press, 1968; Latham et al., 1970b; Lognonné et al., 2009; McGarr et al., 1969), on Mars (Davis, 1993; Teenby and Wookey, 2011), and on asteroids (Ball et al., 2004; Richardson et al., 2004; Walker and Huebner, 2004). Large uncertainties remain when estimating the best value for seismic efficiency, which is the ratio between impact kinetic energy and radiated seismic energy, and values range from 10^{-6} to 10^{-4}, leading to an order-of-magnitude uncertainty in the amplitude of impact-generated seismic waves.

In practice, at teleseismic distances, only LP waves are detected, and the equivalent source of a meteorite impact is a point-force seismic source (Lognonné et al., 2009; McGarr et al., 1969). For low-velocity impacts, the amplitude below the impact cutoff frequency is proportional to the linear momentum of the impactor, while, for larger velocities, the amplitude is increased by up to a factor of 2 due to impact-generated ejecta (Lognonné et al., 2009). These LP source models have been tested successfully with the Apollo seismic data recorded during the controlled impacts of the Saturn V upper stages and the Lunar Modules (see, e.g., the seismic records of the Apollo 17 upper stage, shown in **Figure 15**, with a momentum of about 37 000 kg km s^{-1} (14 487 kg at 2.55 km s^{-1})), and the LP models have been used to estimate the momentum of natural impacts on the Moon (Gagnepain-Beyneix et al., 2006; Gudkova et al., 2011). The amplitude of the largest impacts was estimated by Gagnepain-Beyneix et al. (2006), and models indicate that thes impacts typically release a seismic impulse up to almost 10^9 Ns. This impulse is a factor of 25 larger than the Apollo Saturn V upper-stage impact. Relatively low cut-off frequencies (e.g., up to 1 s) have been found, likely related to the radiation time of the impact-generated shock wave (Gudkova et al., 2011).

The advantage of studying impact-related seismic activity is the reduced number of focal parameters, which are limited to geographical location and time. Moreover, for artificial impacts, these quantities are determined with high precision because the impacts can be radio-tracked. This level of accuracy was achieved for most of the lunar impacts (with the exception of the SIV-B 16 impact, which suffered from a radio-tracking failure), and the precise locations of the artificial impacts have also been confirmed by LRO imaging (e.g., Wolf et al., 2012). As an alternative to artificial impacts, Lognonné et al. (2009) and Yamada et al. (2011) recently proposed monitoring the flashes generated by impacts (e.g., Bouley et al., 2012; Ortiz et al., 2006; Suggs et al., 2008) on the near side and even the far side of the Moon (Mimoun et al., 2012) in order to obtain impact data for future lunar seismic experiments.

Even if the locations of natural impacts are not known, these events are still very useful seismic sources. In the case of the Moon and Earth, and more generally for the terrestrial planets (Le Feuvre and Wieczorek, 2008), the frequency and size of impactors can be obtained using the statistical properties of the Earth-crossing asteroids (e.g., Poveda et al., 1999) and collision rates (e.g., Brown et al., 2002; Shoemaker et al., 1990). Such an approach allows for the modeling of impact frequencies on terrestrial planets that can then be calibrated to the lunar impact-generated seismic noise, using the ALSEP seismic data (Lognonné et al., 2009).

The Moon enables the calibration of impact amplitude and frequency. Typically, impacts have been detected at a rate of about 150 per year, with different detection rates among different Apollo stations resulting from differences in local (site) amplification (Lognonné et al., 2009). The spectrum varies as $A_0 I f^3 / D$, where I is the impulse, f the frequency, D is the epicentral distance, and A_0 depends on the temporal length of the window used for the spectrum computation (Gudkova et al., 2011). For the lunar diffusive case and a window of 5000 s, A_0 is typically found to be 1.1×10^{-12} ms^{-2} Hz$^{-1/2}$, but A_0 can have larger values for shorter windows that are comparable to the rise time of the seismic coda.

As shown by Lognonné et al. (2009), the seismically observed impact rates are well retrieved by a Monte Carlo simulation of impacts constrained by the impact flux rate. **Figure 16(a)** shows the result of such a simulation for different bandwidths, when the spectrum of the signal is as described above and the detection rate is 50 events per year. Over the 7 years of operation, the largest impacts estimated by Gagnepain-Beyneix et al. (2006) and Gudkova et al. (2011) typically released a seismic impulse in the range 10^8–10^9 Ns. This is a factor 10–100 times larger than the Apollo Saturn V upper-stage

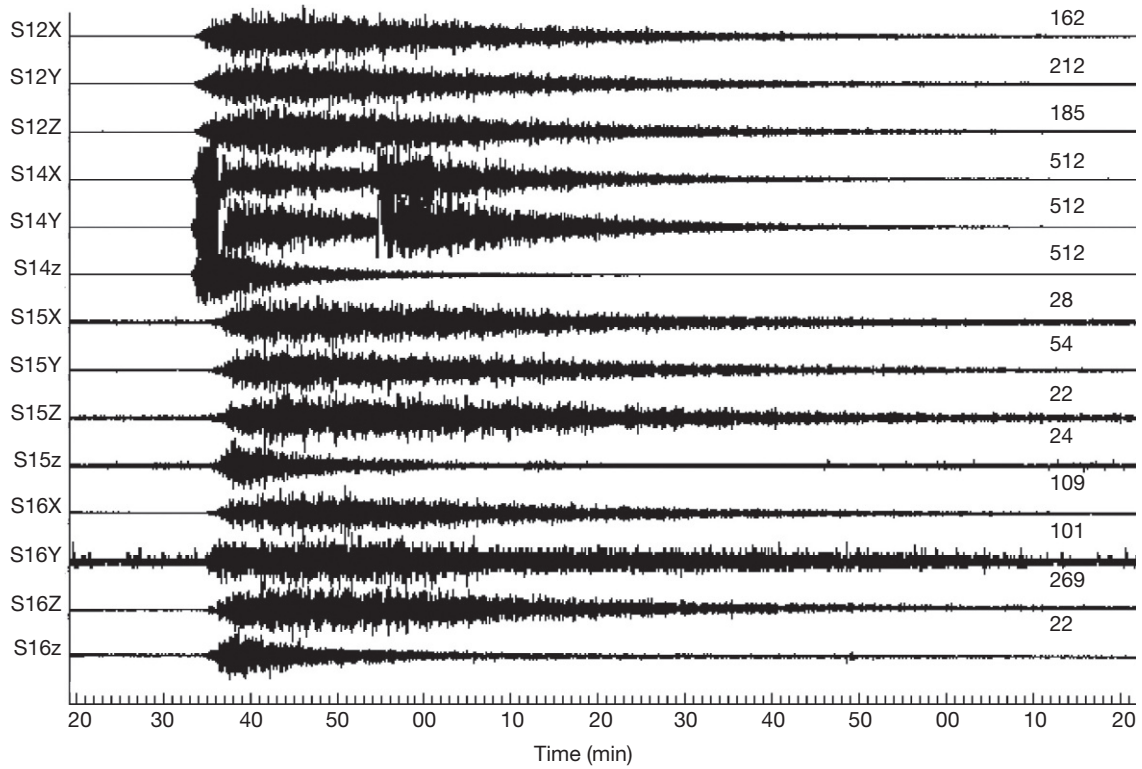


Figure 15 Apollo seismic network records, in digital units, for the impact of the Apollo 17 Saturn V upper stage (Saturn IVB) on the Moon on 10 December 1972 at 4.21° S, 12.31° W (at distances of 338 km from the Apollo 12 station (S12), 157 km from the Apollo 14 station (S14), 1032 km from the Apollo 15 station (S15), and 850 km from the Apollo 16 station (S16)). For each station label, X, Y, and Z represent the LP seismometers, z is for the SP seismometer, and transfer functions are given in **Figure 1**. Amplitudes at the Apollo 14 station, 157 km from impact, are saturated largely due to S waves trapped in the regolith. The first P arrival is typically ten times smaller. Note the 10-db gain change at the middle of the fourth and fifth traces. Reprinted from Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 19.1–19.34. doi:10.1146/annurev.earth.33.092203.122605, with permission of Annual Review of Earth and Planetary Sciences.

impact, and a few hundred times larger than the Lunar Module impact. On the other hand, assuming impacts at 20 km s^{-1}, **Figure 17** shows a typical detection mass threshold of about 100 kg. The detection of smaller impacts requires SP seismometers with sensitivity significantly better than Apollo (e.g., Yamada et al., 2011, 2013).

On planets with an atmosphere, the velocity and the mass of the impactors are reduced during atmospheric entry. This effect can be assessed by integrating the impactor equation in the atmosphere (Chyba et al., 1993; Poveda et al., 1999):

$$m\frac{dv}{dt}=\frac{1}{2}C_D\rho v^2A-mg$$
$$\frac{dm}{dt}=\frac{1}{2}\frac{C_H}{Q}\rho v^3A \qquad [5]$$

where m is the mass of the meteoroid, v is its velocity, ρ is its density, g is the gravitational acceleration, C_D is the drag coefficient (here taken as 1.7, the cylinder value), C_H is the heat transfer coefficient (taken as 0.1), Q is the heat of ablation of the meteoroid (taken as 10^7 J kg^{-1}), A is its cross section, and t is time. **Figure 16(b)** shows the effect of the Mars atmosphere on Mars-impacting meteoroids, with an asteroid density of 2400 kg m^{-3}. The atmospheric shielding effect on Mars is small for the largest events, as it is on the Earth at altitudes of 30 km or higher (ReVelle, 1976) where the pressure is comparable that measured on the martian surface. The model of Le Feuvre and Wieczorek (2008) is used. The entry flux of meteorites is 2.6 times larger than the flux on Earth due to the proximity of the asteroid belt (e.g., Davis, 1993), but the increased flux on Mars is balanced by the impact velocity, which is about half the impact velocity on Earth. The upper number of impacts per year detected by a seismic station on Mars might therefore be comparable to that on the Moon (Davis, 1993), and Martian impacts have been estimated to generate about ten seismic events per year with an amplitude larger than 3×10^{-9} m s^{-2} (**Figure 17**). These seismic events would be associated with impacts of objects having masses of several hundred kilograms, at impact velocities about twice the velocities of lunar impacts.

Most of the differences between impact-related seismic activity on Mars and the Moon are expected to be associated with contrasting crustal seismic attenuation. The attenuation for Mars remains unknown, but it is likely less than the attenuation on Earth. In addition to seismic waves generated by an impact, the interaction of a meteor blast with the martian surface might be a significant source of additional seismic signals through air-coupled Rayleigh waves and direct airwave coupling (Edwards, 2008). Such surface vibrations have been proposed as a preferred source for dust avalanches (Burleigh et al., 2012), rather than pure seismic shaking (Chuang et al., 2007).

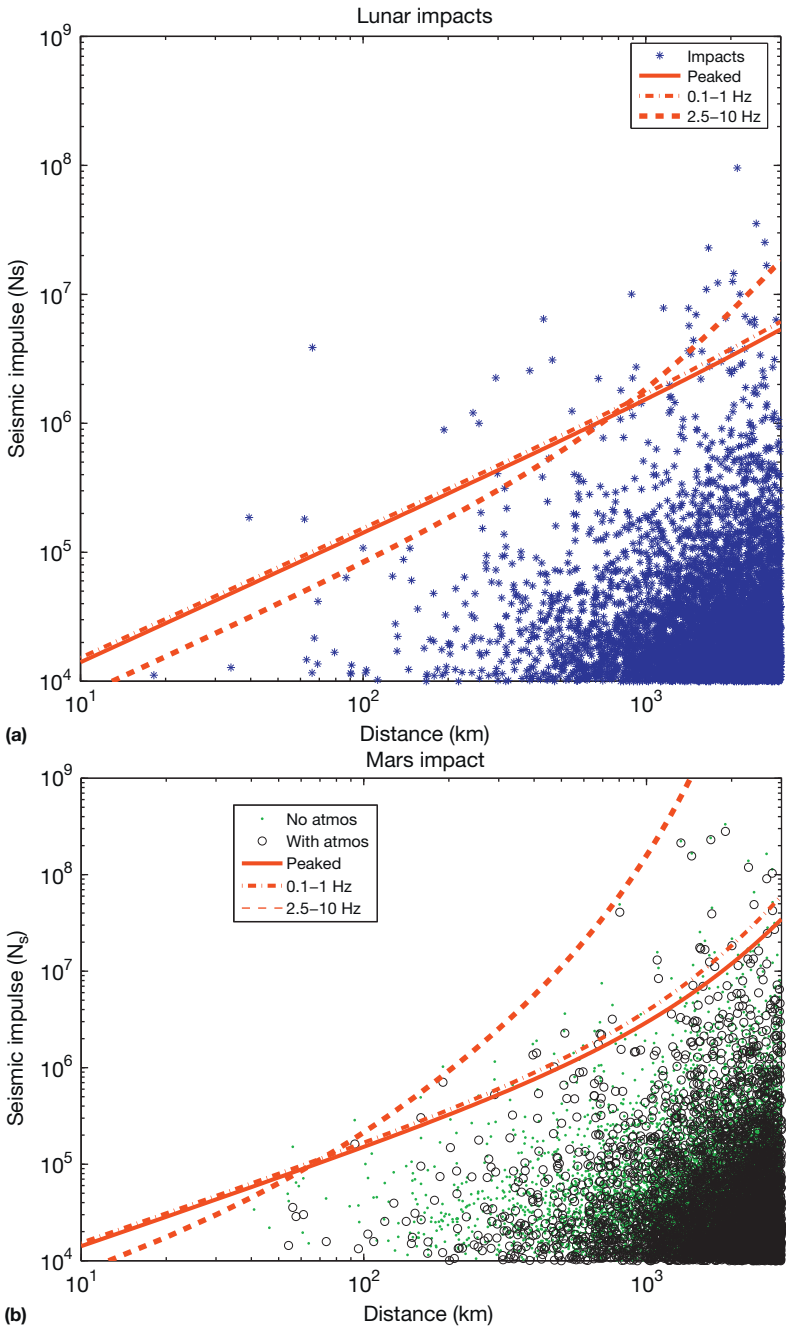


Figure 16 Impact simulation for asteroid impacts on the Moon (a) and Mars (b). The rate of impacts is from Ortiz et al. (2006) and follows the statistical approach of Le Feuvre and Wieczorek (2008), with seismic amplitudes estimated using the approach described by Lognonné et al. (2009). The detection thresholds are based on a signal-to-noise ratio of 3 and flat noise levels of 3×10^{-10} ms^{-2} Hz$^{-1/2}$ in the 0.33–0.75 Hz bandwidth, 5×10^{-10} ms^{-2} Hz$^{-1/2}$ in the 0.1–1 Hz bandwidth, and 10^{-8} ms^{-2} Hz$^{-1/2}$ in the 2.5–10 Hz bandwidth. The first and the last frequency bands are comparable, but not identical, to the Apollo peaked LP and SP detection threshold. For the LP peaked case, this translates into peak-to-zero signal amplitudes of approximately 1.5 DU or 2 mm, where DU $= 0.5 \times 10^{-10}$ m, or ground displacement. The mm value is the amplitude from the Namamura catalog. For the Moon (a), amplitudes have been calibrated with Apollo data and assume a low attenuation ($Q = 5000$) and $v_s = 4.5$ km s^{-1}. For Mars (b), small green dots are impact impulses without atmospheric ablation, while circles represent impulses with atmospheric ablation. Attenuation is corrected by assuming a $Q = 600$ and a shear velocity of $v_s = 4.5$ km s^{-1}, which leads to a corrective factor of $\exp(-\omega t/2Q)$, where $t = D/v_s$. Given a smaller diffusion–scattering than is present on the Moon, larger amplitudes are very likely but not accounted for in these amplitude estimates.

Lastly, we consider small bodies. Currently, the experiment SESAME/CASSE is onboard the Rosetta mission to comet 67P/Churyumov–Gerasimenko (Seidensticker et al., 2007), with a landing expected in 2014. This experiment aims to investigate the outermost surface of the comet by means of active and passive acoustic-wave monitoring in a frequency range from a few hundred hertz to several kilohertz, and the mission may reopen the study of small-body seismology after

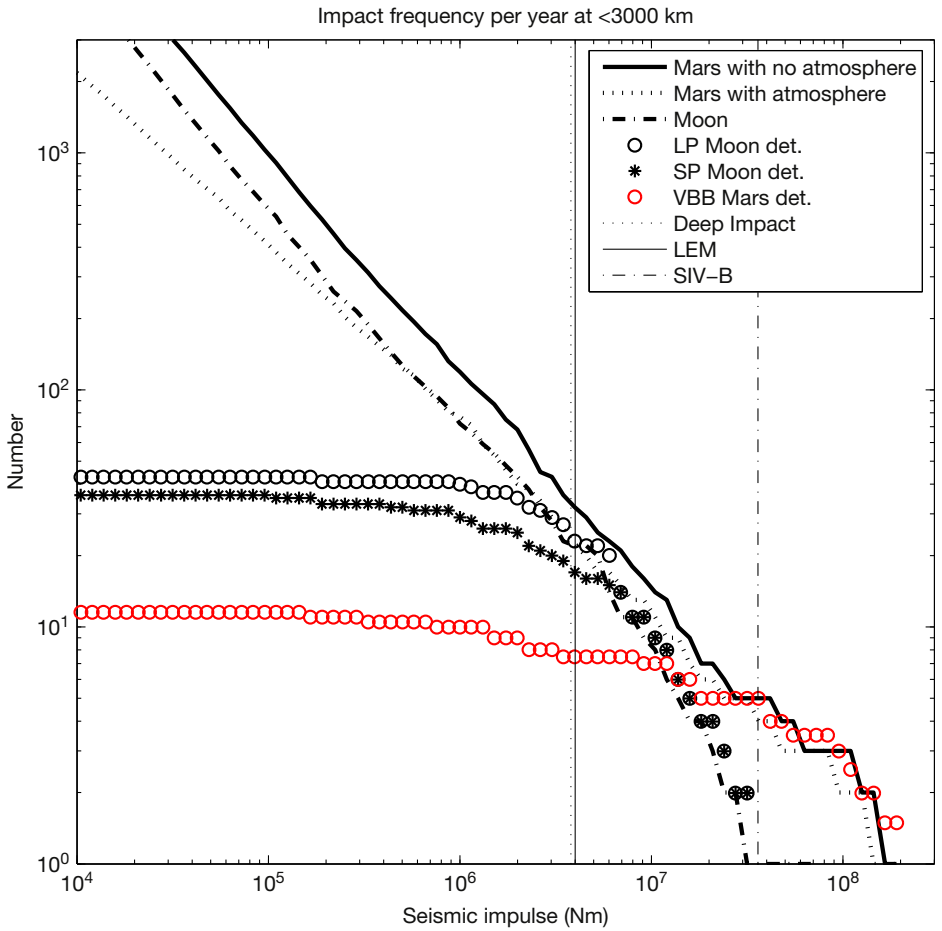


Figure 17 Simulation presented in **Figure 16**. For the Moon and Mars, the detection statistics are the cumulative number of events detected over a distance of 3000 km. The example shown corresponds to a cumulative rate of about 40 detections on the Moon with peak-to-peak amplitudes of 3 DU on the peaked LPZ. This corresponds to the minimum detection rate observed at the Apollo stations, while the maximum rate is about two times larger (see Figure 2 of Lognonné et al., 2009 or **Figure 10** for Apollo 14). Vertical lines provide the impulse values for the Apollo artificial impacts (see Gudkova et al., 2011) and the deep impact performed by the Stardust-NExT (A'Hearn et al., 2005; Schultz et al., 2013). Large uncertainties for Mars remain.

the failure of acoustic sensors onboard the Phobos 1 and 2 landers (Surkov, 1990). Both internal (related to degassing events) and external sources might generate seismic signals on comets, in contrast to asteroids for which only impacts are expected to cause seismicity. Future missions may be targeted toward asteroids, either to study the interior or as a deflection strategy in asteroid mitigation (Ball et al., 2004). Both active seismology (using explosives or impactors) or passive seismology (with natural impacts) have already been proposed (e.g., Walker and Huebner, 2004). Due to their limited size, these bodies have low gravity. As a consequence, minor impacts on asteroids produce ground accelerations higher than the local gravity, and seismic reverberation may have a major effect on the morphology of the asteroid regolith, including the shape and density of small craters (Richardson et al., 2004). The amplitudes of the body's impact-generated P and S seismic waves can be estimated with simple models. Scattering processes will affect body-wave amplitudes, and the irregular asteroid surface will strongly affect the amplitudes of surface waves.

For large impacts, such as the artificial impacts made in the Deep Impact mission (3.8×10^6 Ns by the impact of 370 kg at 10.2 km s^{-1}) (A'Hearn et al., 2005) or those expected in the Don Quijotte mission (4×10^6 Ns by 400 kg at 10 km s^{-1}) (e.g., Ball et al., 2004), the displacement amplitude of body waves can be roughly expressed by:

$$a_p = \frac{F\left(t - \frac{r}{c_p}\right)}{4\pi\rho c_p^2 r} \quad \text{and} \quad a_s = \frac{F\left(t - \frac{r}{c_s}\right)}{4\pi\rho c_s^2 r} \qquad [6]$$

where c_p and c_s are P and S velocities, ρ is bulk density, r is radial distance (Aki and Richards, 2002), and F refers to the point-force source functions (in N). By scaling the impact probability of **Figure 16** by the ratio of the the bodies' surface areas, we see that these impacts, on a 5-km-radius asteroid (i.e., with about 8×10^{-6} less surface area than the moon), will occur naturally at a rate of about one event every 4000–5000 years for an Earth-crossing asteroid flux, and the impacts will generate accelerations in the range of 0.01–1 ms^{-2}, depending on frequency. These accelerations are greater than the local gravity, and seismic reverberations will therefore affect the morphology of the asteroid regolith, including in the shape

and density of small craters (Richardson et al., 2004). As the frequency of impacts is typically a power law function of diameter, with a power law exponent of −5/2 (e.g., Poveda et al., 1999) (and hence a power-law function of mass, with exponent −5/6), accelerations 10^6 times smaller (in the range of 1–100 ng) will occur at a rate 10^5 times higher at 1 AU, and possibly even higher rates will occur at longer distances from the Sun. This translates into rates of hundreds of events per year, which could possibly be recorded by space-qualified SP seismometers. Such natural impacts might provide excellent seismic sources for tomography studies of asteroids (Roark et al., 2010).

10.03.4 Atmospheric Seismology

10.03.4.1 Theoretical Background

Over the last decade, many observations in terrestrial seismology have shown that the coupling between the atmosphere and the solid part of the Earth leads to an exchange of seismic energy between the two subsystems. For example, atmospheric-solid body coupling has been detected in the form of seismic signlas in the ionosphere. As shown by Lognonné et al. (1998b), a small fraction of the seismic energy in the lithosphere escapes into the atmosphere at frequencies higher than 3.7–4.4 mHz. Near the seismic source, the generated acoustic waves propagate as a plume that can reach the high atmosphere after being amplified by the exponential decay of the atmospheric density (Kelley et al., 1985). At greater distances, acoustic waves are primarily generated by Rayleigh surface waves and lead to ionospheric oscillations. These oscillations were first detected in the 1960s, following very large quakes (e.g., Davis and Baker, 1965; Leonard and Barnes, 1965; Weaver et al., 1970; Yuen et al., 1969), and they are now more commonly observed for quakes with magnitudes >7–7.5, using global positioning system (GPS) (Calais and Minster, 1995; Ducic et al., 2003; Garcia et al., 2005a) or Doppler sounders (Artru et al., 2001, 2004).

The excitation of seismic waves by volcanic eruptions (e.g., Kanamori et al., 1994; Lognonné, 2009; Zürn and Widmer, 1996) and the continuous excitation of normal modes (Kobayashi and Nishida, 1998; Suda et al., 1998; Tanimoto et al., 1998), called Earth's hum, have been proposed as other examples of such atmospheric coupling, that time with transfer from atmospheric sources to observation at the surface of the solid Earth.

In the case of Earth's hum, the amplitudes of the seismic signals are small, ranging from 0.3 to 0.5 ng (10^{-11} m s^{-2}) between 2 and 7 mHz, but, nonetheless, the signals are strong enough to allow for the identification of the resonance peaks of the fundamental normal modes. Early studies proposed that turbulences in atmospheric boundary layers (Nishida et al., 2000; Tanimoto, 1999) were strong enough to generate the observed signal. Recent studies show, however, that the source is located over the oceans (Rhie and Romanowicz, 2004, 2006; Tanimoto, 2005) and that infragravity waves over the continental shelves are much more likely seismic sources (Webbs, 2007). In summary, the Earth's hum seems to result from several coupling effects through which energy from the atmosphere is first transferred into oceanic infragravity waves and finally to the Earth's interior. If such a transfer is more efficient than the direct excitation of the normal modes by the atmosphere, what will be the case for a planet without ocean?

This question is addressed regarding Earth's two nearest neighbors, for which comparable energy is transferred into the atmosphere by the Sun. The solar flux driving atmospheric fluid layer circulations is proportional to $(1-A)/D^2$, where D is the distance to the Sun and A is the bond albedo (de Pater and Lissauer, 2001). The relative ratios of the Venus and Mars fluxes to the Earth flux are 0.67 and 0.51, respectively, and these flux ratios demonstrate the need to quantify the seismic energy-conversion efficiency for planets without oceans.

As we will see, the previously discussed topics might offer tremendous opportunities for new seismological discoveries. For Mars, continuous excitation of free oscillations, if strong enough, could enable the determination of fundamental mode frequencies, even if no large quake is directly detected. For Venus, remote detection in the atmosphere might provide the opportunity to perform seismic measurements, avoiding the major challenges associated with a long-lived lander able to withstand the planet's 500 °C surface temperature. More generally, any planetary body with an atmosphere will exhibit acoustic–seismic coupling between the atmospheric and interior parts of the planet. Other bodies of particular interest in this regard include Titan and the giant planets, for which no discontinuity exists between the interior and the atmosphere. Therefore, in this section we explore atmospheric coupling on Mars and Venus, focusing on the expected amplitude of signals. We then provide a short review of the status of giant planet seismology, for which the first seismic observations are yet to be performed.

We start by reviewing the main properties related to surface-atmosphere seismic coupling. The density and sonic speed for the lowest 150 km of the Earth, Venus, and Mars atmospheres are given in **Table 2** and **Figure 18**. The theory of interior-atmospheric coupling must first take into account the fact that a fraction of the seismic-wave energy is transmitted into the atmosphere when the seismic wave reaches the surface of the solid (or for the Earth, liquid) planet. For a vertically incident

Table 2 Atmospheric coupling parameters for Venus, Earth, and Mars

	Crust	*Earth*	*Mars*	*Venus*
Surface sound speed (m s^{-1})	5800	340	214	426
Surface density (kg m^{-3})	2600	1.225	0.0175	65
Acoustic impedance ratio/crust	1	2.76e−5	2.48e−7	1.83e−3
Energy transfer $\left(E=\frac{Q}{\pi}\frac{\rho_{air}c_{air}}{\rho_{int}c_{int}}\right)$ $(Q=100)$		8.8e−4	7.9e−7	5.8e−2
High-Q atmospheric resonances (mHz)		3.7	2.05	3.10; 4.15

The atmospheric model used for Earth is the US Reference Atmospheric model. For Mars, the atmospheric model is the mean model from a GCM run of Forget et al. (1999).

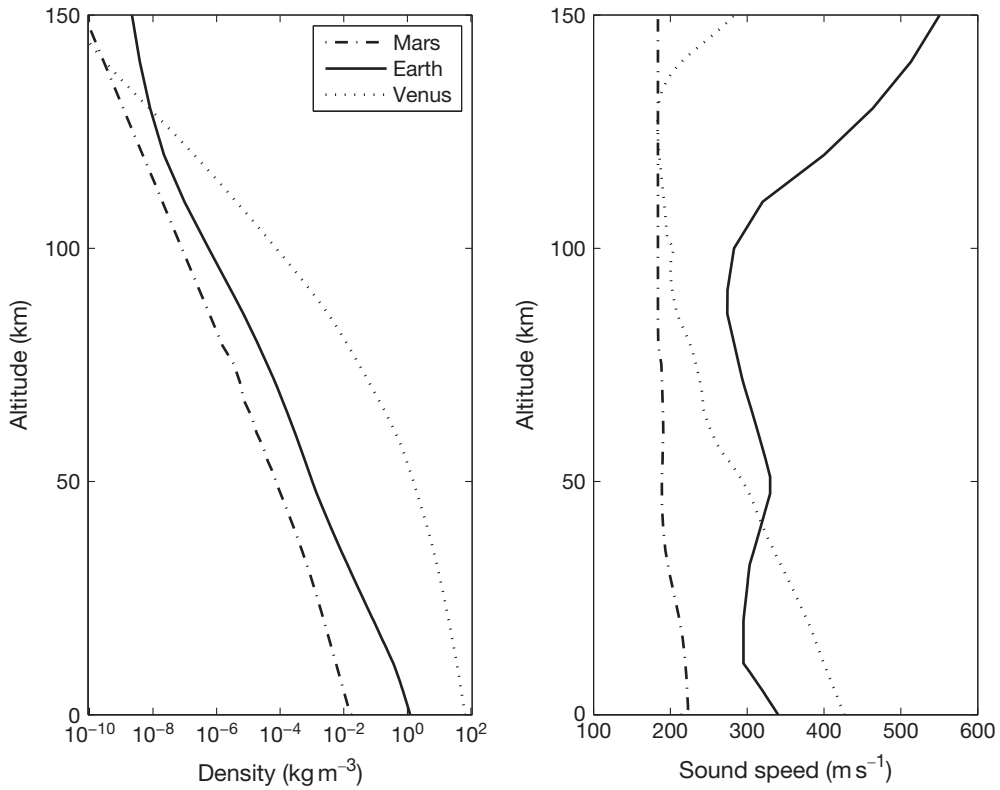


Figure 18 Atmospheric models of the Earth, Mars, and Venus used for the computation of atmospheric coupling. Note that these models are sensitive to the local time and latitude, as well as to the solar environment. The US standard atmospheric model (1976) is used for the Earth, whereas the models of Forget et al. (1999) and Hunten et al. (1983) are used for Mars and Venus, respectively.

acoustic wave, the amplitudes of transmission (t) and reflection (r) coefficients from the interior to the atmosphere are given by:

$$t = \frac{2\rho_{\text{int}}c_{\text{int}}}{\rho_{\text{air}}c_{\text{air}} + \rho_{\text{int}}c_{\text{int}}} \approx 2$$
$$r = \frac{\rho_{\text{int}}c_{\text{int}} - \rho_{\text{air}}c_{\text{air}}}{\rho_{\text{int}}c_{\text{int}} + \rho_{\text{air}}c_{\text{air}}} \approx 1 \qquad [7]$$

while the energy transmission (T) and reflection (R) coefficients are

$$T = \frac{\rho_{\text{air}}c_{\text{air}}}{\rho_{\text{int}}c_{\text{int}}}t^2$$
$$R = r^2 \qquad [8]$$

and $R+T=1$. The initial amplitude of the waves in the atmosphere is therefore twice the vertical amplitude of the waves (or the amplitude of the ground, due to continuity), while the relative transmitted energy is four times the acoustic impedance ratio. Surface waves of angular order ℓ have a horizontal wavelength of $\lambda = 2\pi a/(\ell+1/2)$ and will bounce on the surface once per cycle, with amplitude decreasing each time by $e^{-\pi/Q}$ due to attenuation in the solid part, where Q is the quality coefficient of the mode. Thus, during these successive bounces, the portion of the energy transmitted into the atmosphere will be $E = e(2Q/\pi)(\rho_{\text{air}}c_{\text{air}}/\rho_{\text{int}}c_{\text{int}})$, where e is the partition ratio between the energy in vertical displacement and the total energy, the latter being typically around 0.5. Such energy transmission, given in **Table 2**, is approximately found in the high ℓ limit of **Figure 19**.

In the atmosphere, in the non-viscous case, the seismic wave equation is given by:

$$\rho_0 \frac{\partial \vec{v}}{\partial t} = -\vec{\nabla} p + \rho_1 \vec{g}_0 + \rho_0 \vec{g}_1$$
$$\frac{\partial \rho_1}{\partial t} + \text{div}\left(\rho_0 \vec{v}\right) = 0$$
$$\text{div}\left(\vec{g}_1\right) = -4\pi G\rho_1$$
$$\frac{\partial p}{\partial t} = -\gamma p_0 \text{div}\left(\vec{v}\right) - \rho_0 \vec{g}_0 \cdot \vec{v} \qquad [9]$$

where ρ_0, ρ_1, and p are the equilibrium density, perturbation in density, and pressure, respectively, and where $\vec{v}$, $\vec{g}_0$, and $\vec{g}_1$ are the velocity, equilibrium gravity, and perturbation in gravity associated with the mass redistribution. p_0, γ, and G are the equilibrium pressure, the adiabatic index, and the gravity constant. The upward propagation of the waves is modeled by renormalizing the amplitudes in order to account for the exponential decay of the density. Quasianalytical expressions can be obtained when the atmospheric structure has a vertical scale much larger than the wavelength of the wave and when pure upward propagation is assumed. Gough (1986) shows that under these conditions eqn [9] leads to

$$\left(\frac{\partial^2}{\partial t^2} + \omega_c^2\right)\frac{\partial^2 \psi}{\partial t^2} - c^2 \frac{\partial^2}{\partial t^2}\nabla^2 \psi = 0 \qquad [10]$$

where the cutoff frequency is given in first approximation by $\omega_c = c/2H_\rho$ and where $\psi = c^2\sqrt{\rho_0}\text{div}\left(\vec{v}\right)$ is related to the

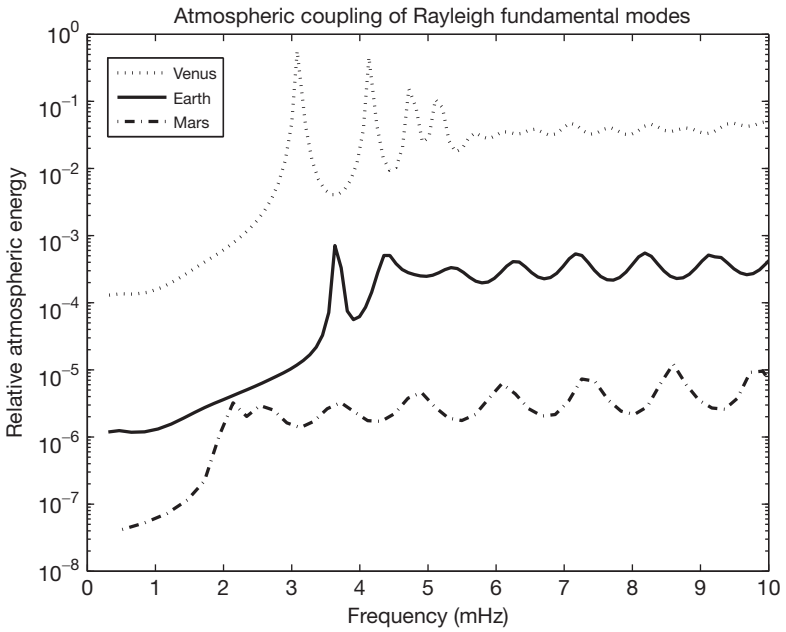


Figure 19 Fraction of the surface-wave energy in the Venus, Earth, and Mars atmospheres for Rayleigh surface waves. Only the first peaks are due to atmospheric resonances. Note that the amplitudes on Mars and Earth are comparable at low frequency (2–3 mHz) due to differences in the atmospheric resonance frequency.

square root of the acoustic energy $c\gamma_0 p_0 \mathrm{div}^2\left(\vec{v}\right)$, with $c^2 = \gamma_0 p_0/\rho_0$ being the squared speed of sound. Here, H_ρ is the density height scale given by $H_\rho = -(\mathrm{d}\,\mathrm{Ln}\,\rho/\mathrm{d}r)^{-1}$. The plane wave solution to eqn [10] obeys the dispersion equation:

$$\omega^2 = \omega_c^2 + k^2 c^2 \qquad [11]$$

Waves with frequencies lower than the cutoff frequency have negative k^2 and exponentially decaying energy in the atmosphere. Waves with frequencies higher than the cutoff frequency can propagate upward. This energy is not trapped, and the waves progressively lose energy during their upward propagation through viscous and nonlinear processes. Note that other parameterizations are possible, leading to slightly different expressions for the dispersion equation and cutoff frequencies (Beer, 1974; Mosser, 1995). The trapping of free oscillations, which can be used to model the waves with summation techniques, can be modeled when the full system (i.e., the solid and atmospheric parts of a planet) is considered, and when a radiating boundary condition (Lognonné et al., 1998b) is assumed at the top of the atmosphere. We provide the results of forward modeling based on this theory for Earth, Mars, Venus, and Jupiter.

For planets with solid parts, the normal modes of the system can be computed for spheroidal solid modes, atmospheric acoustic modes, and atmospheric gravity modes. The summation of these normal modes then allows the computation of either seismograms for the solid surface and for the atmospheric perturbations or barograms for the atmosphere as a function of altitude. More details and the theoretical background can be found in Lognonné et al. (1998b) and Lognonné and Clévédé (2002). An extension of these coupling models to the viscous case is given by Artru et al. (2001). Figure 19 illustrates the efficiency of coupling for Venus, Mars, and Earth, and it shows the fraction of energy of the fundamental Rayleigh modes in the atmosphere when the latter are computed with consideration given to atmospheric coupling. Resonance effects appear with a peak in the atmospheric energy; they are observed at frequencies associated with the fundamental and overtones of the atmospheric waves guide and are dependent upon the atmospheric model used.

At these frequencies, a much larger fraction of the seismic waves is transferred into the atmosphere. On Earth such resonance frequencies between the atmosphere and interior exist near 3.7 and 4.44 mHz and are preferential windows for studying atmospheric-interior coupling. Large atmospheric eruptions, such as those of El Chichon in 1982 and Pinatubo in 1991, led to selective excitation of Rayleigh surface waves at these frequencies (Kanamori and Mori, 1992; Widmer and Zürn, 1992; Zürn and Widmer, 1996). The fundamental resonance frequency on Mars is found at about 2.05 mHz, while the fundamental frequency and the three next overtones on Venus are found at 3.10, 4.15, 4.75, and 5.15 mHz. On Venus the two first overtones are strongly trapped while only the fundamental branch is trapped on the Earth and Mars.

Figure 20 shows the Rayleigh normal modes for angular orders smaller than 50. Below the atmospheric fundamental frequency, amplitudes decay with altitude, and full trapping of the modes is observed. At higher frequencies, oscillations appear as a consequence of the upward propagation associated with energy leakage, and the normal modes also have significant imaginary amplitudes in the atmosphere. Due to its smaller size and to a low resonance frequency (associated with the cold atmospheric temperature), this transition for Mars is found at $\ell = 8$, while it is observed for $\ell = 28$ and 22 for the Earth and Venus, respectively. The amplitudes are also proportional to the ratio of acoustic impedance between the ground and atmosphere.

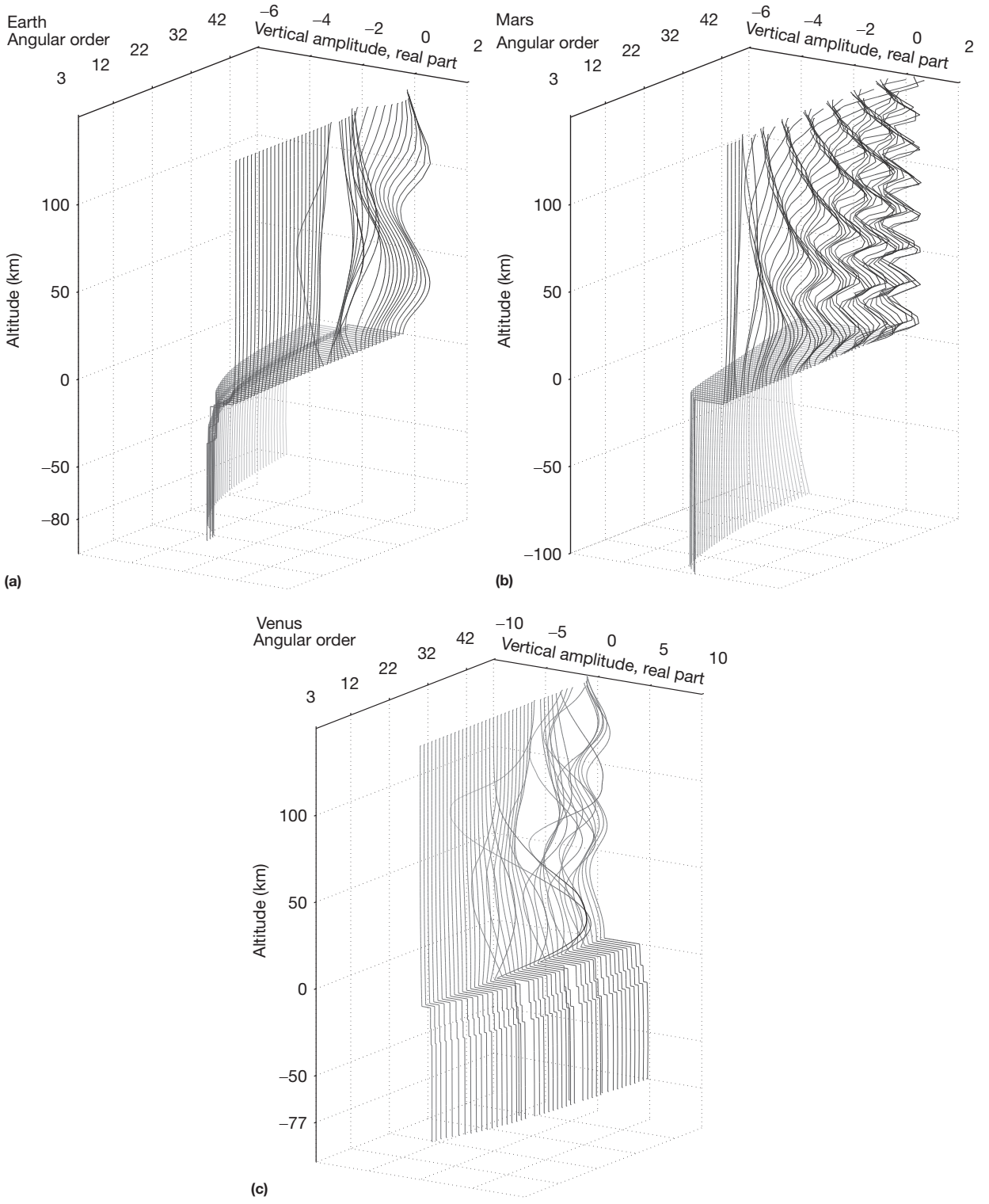


Figure 20 Rayleigh normal modes for Earth (a), Mars (b), and Venus (c). The amplitudes shown are the vertical displacement multiplied by the square root of density. Amplitudes are multiplied by 5 for the Earth and by 50 for Mars. For Venus, amplitudes are not amplified. Amplitudes are shown over a vertical scale from a depth of 100 km to an altitude of 150 km. Note that, in all cases, the amplitudes decay exponentially when close to the surface for the low angular orders below the cutoff frequency, with the single oscillation at the resonance comparable to the amplitude of the atmospheric fundamental modes and the oscillating amplitudes above the resonance frequency.

10.03.4.2 Mars Hum and Martian Atmospheric Sources

We now address the possibility of the continuous excitation of normal modes by the martian atmosphere. Kobayashi and Nishida (1998) originally proposed amplitudes on Mars to be a factor of 2–3 smaller than those on Earth, while amplitudes on Venus were expected to be comparable with Earth amplitudes. These early studies primarily considered excitation by turbulence in the boundary layer, however. (See more on the theory and assumptions in Tanimoto (1999, 2001).) This theory was later criticized by Webb (2007), who showed that it yields an overestimation of the normal mode amplitudes by several orders of magnitude.

Can we still imagine that normal modes might be excited with potentially observable amplitudes? Instead of turbulence, let us consider nonturbulent large-scale winds and atmospheric circulation. As they generate wind Reynolds stresses, they act as a potential seismic source. As shown by Lognonné et al. (1994), the density of pressure glut can then be expressed as

$$m_{ij} = \left(p + \kappa \nabla \cdot \vec{u}\right)\delta_{ij} + \rho v_i v_j \qquad [12]$$

where p is the atmospheric pressure, v_i is the ith wind velocity component, ρ is the density, κ is the bulk modulus of the atmosphere, and $\vec{u}$ is the displacement field of the normal modes in the atmosphere. In the isotropic case, when wind is neglected, the pressure is much greater than the Hook pressure of normal modes, and we have $m_{ij} = p d_{ij}$. The amplitude of a normal mode with angular order ℓ and eigenfrequency ω_ℓ can then be expressed with a similar formalism, as in Lognonné et al. (1994), which leads to

$$a_\ell^m(t) = \int_{-\infty}^{t} \mathrm{d}\tau \, \mathrm{e}^{\mathrm{i}\omega_\ell(t-\tau)} \iint \mathrm{d}\Sigma \, \overline{Y}_\ell^m(\theta, \phi) \int_0^{+\infty} \mathrm{d}z A_\ell(z) \frac{Mp(z, \theta, \phi, t)}{R\rho_0(z)} \qquad [13]$$

where z, θ, ϕ are the altitude, latitude, and longitude, respectively; $a_\ell^m(t)$ is the acceleration of a given normal mode at a time t; $A_\ell(Z)$ is an excitation term that depends on the atmospheric-solid planet coupling, related to the divergence of the normal-mode eigenfunction times, the amplitude of the mode on the recording direction; $\overline{Y}_\ell^m$ is the complex conjugate of spherical harmonic of degree ℓ and order m; p is the atmospheric pressure; R is the thermodynamical constant; ρ_0 is the atmospheric density; and M is the mean atmospheric molar mass (Note that $Mp/R\rho_0$ has the dimension of temperature.). Values of the excitation terms, $A_\ell(Z)$, at the surface of the planets (z=0), are shown in **Figure 21** for the fundamental normal modes of Earth and Mars. On Earth, at an angular order up to 28 or 29, acoustic energy is trapped and therefore not radiated. On the other hand, acoustic trapping on Mars occurs at an angular order below 10 for all realistic atmospheric models, leading to a dynamic coupling of the planet's normal modes and atmospheric circulation that is quite efficient for low-order angular modes. Thus, the coupling coefficient for Mars is comparable to the coefficient for Earth at angular orders of about 10, but, as a consequence of the lower atmospheric density on Mars, it is typically a factor of 10 smaller than Earth's coupling coefficient for angular orders >20. Temperature fluctuations and winds are significantly larger on Mars, however.

Global circulation models (Forget et al., 1999) can be used to obtain more precise estimates of the continuous excitation of normal modes through a computation of the pressure glut. **Figure 22** shows the right-hand side of eqn [13], which corresponds to the excitation term related to the pressure field at a given local time and for a classical global circulation model (GCM) just before time-integration and multiplication by $\mathrm{e}^{\mathrm{i}\omega_\ell(t-\tau)}$. Large excitations are observed in this model, including excitations along atmospheric fronts in the southern hemisphere. After time-integration, such models provide the

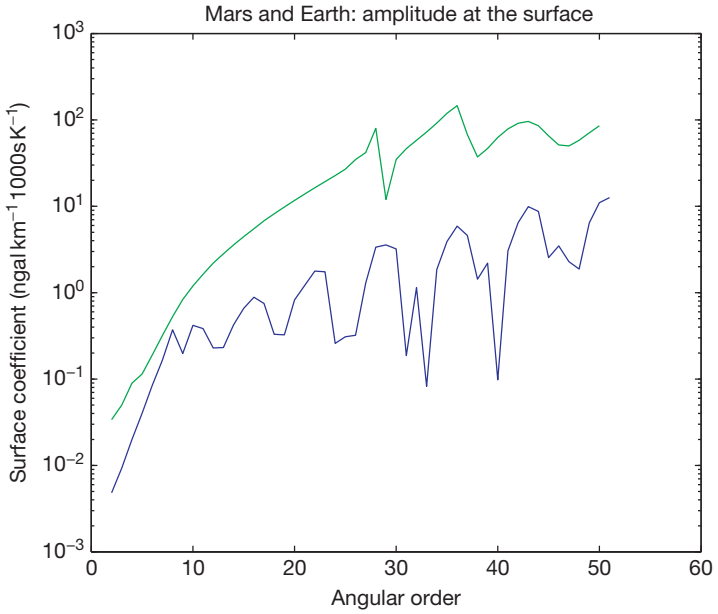


Figure 21 Excitation term $A_\ell(Z)$ versus angular order for Earth and Mars (eqn [13]) at the surface of the planet (z=0). Note that, at low angular order, the excitation terms for the two planets are comparable. Due to the high temperature variations, the martian hum might therefore be more efficient for these low order terms than on Earth.

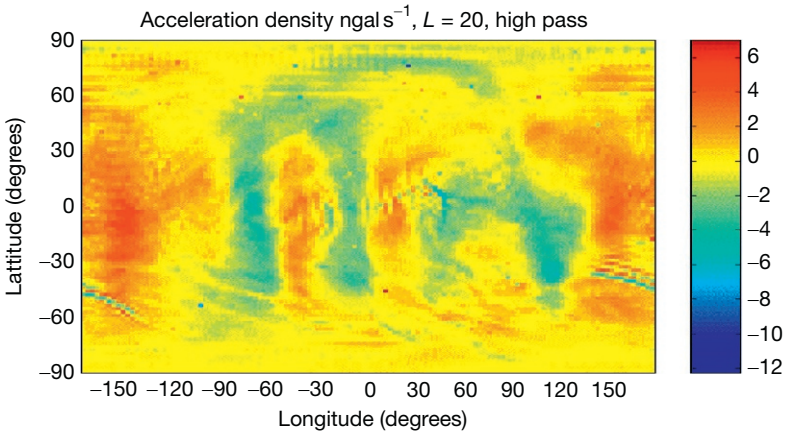


Figure 22 Acceleration field of a martian GCM model at a given time on the planet, obtained for a 30-day run of the GCM (Forget et al., 1999), corresponding to the term $a_\ell^m(t)$ in eqn [13]. The field is computed for the spheroidal mode ${}_0S_{20}$ and includes the vertical, last integration of eqn [13], but not the Legendre transformation or the time integration. Note the presence of the atmospheric front in the southern hemisphere, which provides highly localized excitation of the normal modes, as well as the large-scale pressure field associated with the daily variation.

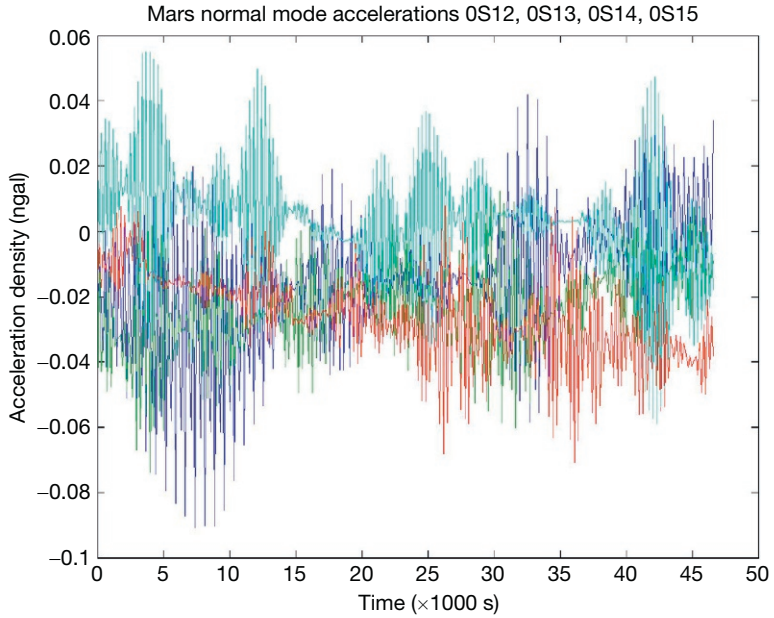


Figure 23 Simulations of the permanent excitation of normal modes for the spheroidal modes ${}_0S_{12–15}$ on Mars, with a GCM model, during a 2-week period. An interval of about 14 h is shown, illustrating the typical amplitude variations for these modes, with typical periods from 400 to 300 s. Increasing angular orders are coded blue, green, red, and cyan, respectively. Observed amplitudes on Earth below an angular order of 20 are 0.2 ngals.

acceleration amplitudes of the modes (Figure 23). As mentioned above, these estimates probably represent minimum values, because GCMs do not model short-scale circulations in the atmospheric boundary layer. Yet, the results still show that a small excitation is achieved by atmospheric global circulation. Moreover, the global circulation on Mars is more coherent than the circulation on Earth from one day to the next. So, if long-duration measurements become available, increasing signal strength through stacking methods may be possible. Even though the search for continuous excitations will be challenging, due to the extremely low amplitudes and the high noise associated with the surface installation of seismometers on Mars, it can open exciting opportunities for the detection of low-angular-order normal modes, and it will benefit from long-lived seismic monitoring and the stacking methods that monitoring would enable.

We now briefly consider the effect of the atmosphere over shorter periods. Atmospheric wind-induced turbulences are known to be a source of SP seismic noise in the frequency range of 0.1–10 Hz (e.g., Steeples et al., 1997). Recent studies in terrestrial seismology using cross-correlation techniques have shown that the shallow and middle crust can be imaged by extracting Green's function for Rayliegh waves from long sequences of ambient seismic noise generated by atmospheric sources. The same technique has been employed on the Moon at much shorter periods, using thermally generated noise, for

the investigation of the near-subsurface (Larose et al., 2005; Sens-Schönfelder and Larose, 2010; Tanimoto et al., 2008). Such a technique was also applied in California on short period (6–20 s) surface waves (Shapiro and Campillo, 2004; Shapiro et al., 2005), and the technique can be applied to shorter-wavelength surface waves as well, as it is sensitive to the uppermost kilometers of the crust. Martian atmospheric turbulence, especially the localized dust devils, can probably generate surface waves comparable to the waves generated on Earth. Passive short period surface wave observations, possibly from multiple sensors, may provide an alternative to future active seismic experiments for studying the martian subsurface and the associated water reservoirs. This would obviate the technological and safety challenge of carrying explosive seismic sources to Mars, as was done on the Moon for the Apollo seismic experiments.

10.03.4.3 Venus Atmospheric Seismology

As noted earlier, ionospheric disturbances associated with acoustic waves generated by Rayleigh surface waves have been observed in connection with very large ($M_w > 7$) earthquakes (e.g., Artru et al., 2004; Ducic et al., 2003; Garcia et al., 2005a), as well as earthquakes of lesser magnitude. Calais and Minster (1995) reported ionospheric perturbations on GPS data following the Northridge ($M_w = 6.7$) earthquake, and observations have demonstrated the acoustic-coupling effects of earthquakes with magnitudes as low as 5.9, with reported thermospheric perturbations of about 300 K between 300 and 400 km of altitude (Kelley et al., 1985). For a short review of ionospheric postseismic perturbation, see Lognonné et al. (2006). The detection of atmospheric signals associated with quakes might therefore be an interesting alternative to seismometer use, especially if measurements can be performed by orbiting spacecraft or Earth-based telescopes. However, Venus is the only planet where possibly weaker seismic activity is counterbalanced by stronger atmospheric coupling, enabling seismic studies using this alternative strategy.

Despite claims of microseismic signal detection during the operational life of the Venera landers (Ksanfomaliti et al., 1982), the establishment of a long-term seismic station on Venus will be a huge challenge (Stofan et al., 1993), and an orbit-based seismic pathfinder survey is therefore an elegant, low-cost option. We can estimate the amplitude of the atmospheric signals expected to accompany large quakes on Venus. The amplitudes of seismic waves in the Venusian atmosphere are related to the acoustic jump at the surface and the associated transmission coefficient, as well as to altitude-related amplification. At the Venusian surface, the atmospheric pressure is about 90 bars, the density is about 65 kg m^{-3}, acoustic velocities are slightly higher (425 m s^{-1}) than Earth velocities, and ground coupling (ρc) is about 60 times greater. At an altitude of 50 km in Venus' atmosphere, the pressure is comparable to Earth's surface pressure, but the density is almost two orders of magnitude lower than the density at the Venusian surface. These two effects mean that, for a given altitude and quake moment magnitude, atmospheric signals are expected to be about 600 times greater on Venus than on Earth, making orbit-based detection a viable possibility.

The first possibly detectable signal might be the atmospheric and ionospheric perturbation in the vicinity of a Venusian quake. Such a signal will consist of a plume that generates temperature and possibly brightness-albedo, as well as atmospheric–ionospheric oscillations. The near-source characteristics of a generated acoustic plume were first studied by Garcia et al. (2005b), who showed that either adiabatic oscillations corresponding to the acoustic waves or nonadiabatic deposition could be generated. For $M_w = 6$, adiabatic temperature oscillations of about 100 K and postseismic heating of about 10 K were also found (**Figure 24**). These perturbations take place very close to the source, however, within a distance comparable to the event depth, and the observation of such effects will therefore require a global, high-resolution survey that might only be achieved by spacecraft near Venus. The detection of signals related to an acoustic plume was attempted with the Virtis experiment (Drossart et al., 2004), onboard the European Space Agency's (ESA) Venus Express spacecraft, in orbit around Venus without success.

In addition to these signals in the vicinity of the quakes, infrasound and surface waves generated by the quake will also produce atmospheric signals, which can be detected at larger distances. These waves can be computed more precisely using the theory developed by Lognonné et al. (1998b), and, in the cases of both Earth and Venus, the differences in atmospheric coupling can be analyzed via the spherical PREM (Dziewonski and Anderson, 1981), with Venus' interior pressures employed in the PREM applied to that planet.

A first, impressive effect of atmospheric coupling is the strong perturbation of the quality factor of the interior's fundamental normal modes (**Figure 25**). The presence of the atmosphere decreases the attenuation coefficient of Rayleigh modes, Q, due to the escape of energy. For Earth, this effect is comparable to or smaller than the error in the determination of Q. For example, the Earth normal mode ${}_0S_{98}$ is close to 10 mHz and has a Q of 115 with the atmosphere and a Q of 118 without the atmosphere, implying an atmosphere-related Q-decrease of only 2.5%. On Venus, we get a Q of 108 for the mode ${}_0S_{93}$ with the atmosphere and a Q of 121 without the atmosphere, corresponding to a decrease of more than 10%. The difference in atmosphere-related Q-decreases means that about 10% of the energy of the Rayleigh waves is dissipated in the Venus atmosphere, confirming the importance of the coupling that occurs in it.

A comparison of synthetic seismograms for the two planets allows further examination of coupling effects. **Figure 26** shows the atmospheric oscillations for Earth and Venus at a 150-km altitude for the same epicentral distance and seismic source. We find amplitudes about 100 times larger on Venus than on Earth, reaching 0.3 m s^{-1} at these periods and corresponding to wavelengths longer than 300 km. Amplitudes will be about one order of magnitude larger at 20 s but are much smaller, of course, for low-order angular modes. **Figure 27** shows the amplitude for the fundamental Rayleigh mode in the latter case, as well as for the fundamental and two first overtones of the atmospheric waveguide. For a quake releasing a moment of 10^{18} Nm, we get individual mode amplitudes of a few cm s^{-1} at 150-km altitude and angular orders >30. At 120 km, however, amplitudes do not reach 1 mm s^{-1}.

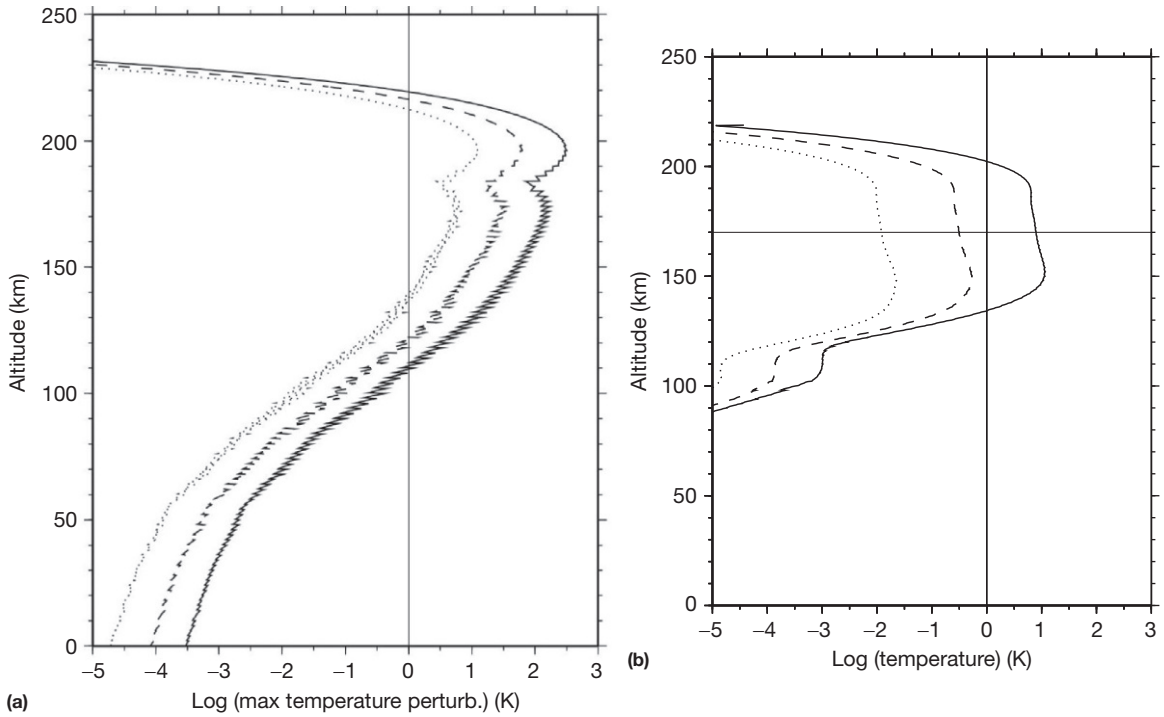


Figure 24 Logarithm of (a) the maximum adiabatic temperature perturbation (in K), as a function of altitude (in km) for earthquake moment magnitudes $M_w = 5$ (dotted line), $M_w = 5.5$ (dashed line), and $M_w = 6$ (plain line). In (a), the signals have periods corresponding to the acoustic waves, and the vertical bar represents a 1-K amplitude. In (b), the figure is the same but for the nonadiabatic heating associated with acoustic energy deposition in the atmosphere, which remains for a longer time before atmospheric thermal diffusion. The additional horizontal line represents the minimum altitude at which the signal is maintained for more than 200 s. At higher altitudes, the temperature anomalies remain for about 4 min. Reprinted from Garcia R, Lognonné P, and Bonnin X (2005) Detecting atmospheric perturbations produced by Venus quakes. *Geophysical Research Letters* 32: L16205; Garcia R, Crespon F, Ducic V, and Lognonné P (2005) 3D ionospheric tomography of post-seismic perturbations produced by the Denali earthquake from GPS data. *Geophysical Journal International* 163: 1049–1064, with permission of the Geophysical Research Letters.

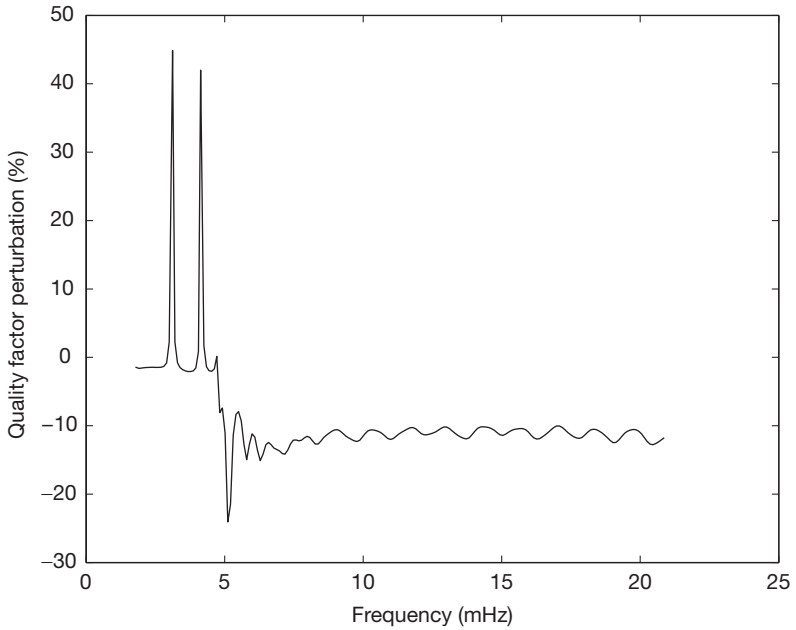


Figure 25 Perturbation of the quality factor for the spheroidal fundamental normal modes. The two peaks are associated with resonances at which significant energy is transferred into the atmosphere, instead of within the solid planet. At these two frequencies, the surface waves have a significant part of their energy in the atmosphere, where attenuation processes are smaller than the attenuation processes in the solid part of the planet. This distribution of wave energy leads to reduced attenuation. The resonance is related to the crossing of the Rayleigh dispersion branch with the fundamental acoustic branch. For the other frequencies, and especially those larger than 5 mHz, modes lose a significant fraction of their energy in the atmosphere, as all the energy transferred across the solid-atmosphere discontinuity propagates upward and dissipates at high altitudes through viscous and nonlinear processes.

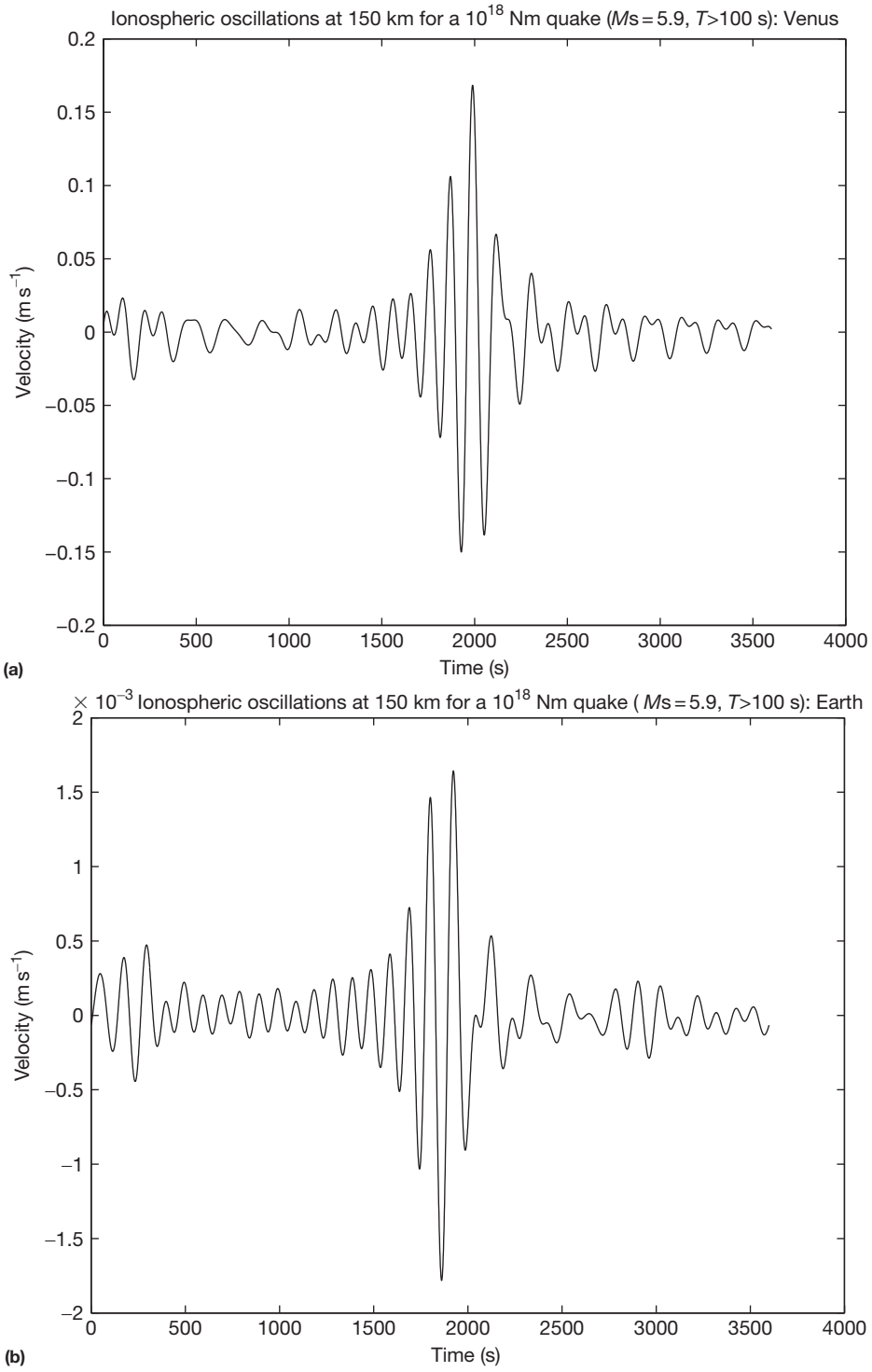

Ionospheric oscillations at 150 km for a 10^18 Nm quake (Ms = 5.9, T>100 s): Venus
Velocity (m s^-1)
Time (s)
(a)
× 10^-3 Ionospheric oscillations at 150 km for a 10^18 Nm quake (Ms = 5.9, T>100 s): Earth
Velocity (m s^-1)
Time (s)
(b)

Figure 26 LP vertical atmospheric oscillations for a 10^{18} Nm quake ($M_w = 5.9$) and for a period longer than 100 s on Venus (a) and on Earth (b). Due to the difference in the acoustic coupling at the ground, at an altitude of 150 km, ionospheric signals on Venus are about 100 times stronger than the signals produced by an earthquake of the same magnitude. This is about 1.3 M_w magnitude. Ionospheric velocity oscillations are about 0.3 m s^{-1} peak-to-peak at these periods, corresponding to wavelengths larger than 300 km. Thus, they will be about one order of magnitude larger at 20 s.

The presence of seismic signals in the Venus atmosphere open the possibilities of remote detection, either from Earth or from Venus orbit. This type of detection can be accomplished using remote sensing techniques of the neutral atmosphere, as done in Jupiter seismology (see **Section 10.03.4.4** below), or by way of ionospheric sounding, as done in Earth ionospheric seismology (e.g., Lognonné, 2009). In regards to neutral-atmosphere remote sensing, the detection of oscillations

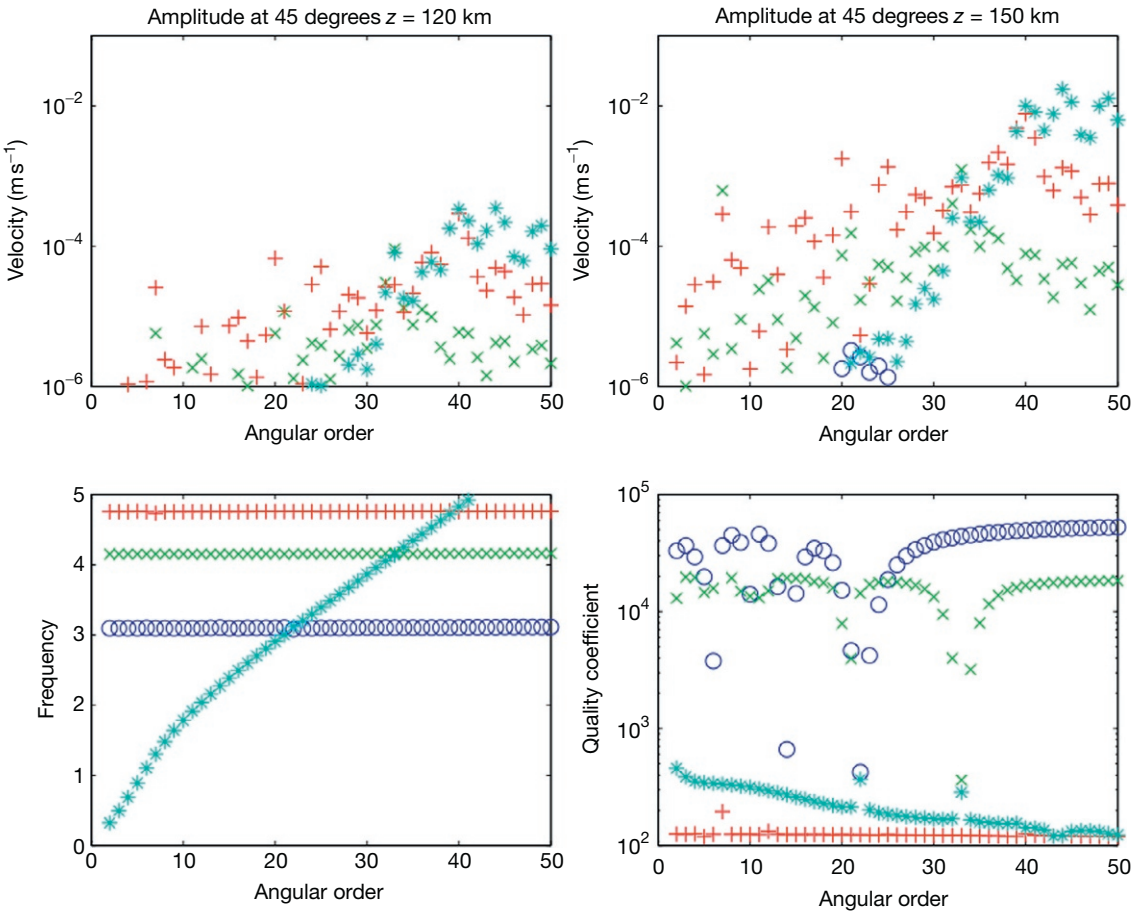

Amplitude at 45 degrees z = 120 km
Amplitude at 45 degrees z = 150 km
Velocity (m s−1)
Angular order
Frequency
Quality coefficient

Figure 27 Free oscillation amplitudes (vertical velocities in m s^{-1}), with respect to angular orders, for a quake with a seismic moment of 10^{18} Nm ($M_w = 5.9$). All amplitudes are expected to be multiplied by the spherical harmonics values at the locations of their respective observations. Sky blue stars represent the Rayleigh fundamental modes. Dark blue circles represent the fundamental acoustic modes. The green × and red + represent the first and second acoustic modes, respectively. For angular orders below 20, the order of the branches is as follows: Rayleigh fundamental modes, fundamental mode, first acoustic overtones, second acoustic overtones. In the intraplate-cooling hypothesis, such quakes are expected at a rate of 2 per month (quakes with moments and therefore amplitudes ten times larger are expected at a rate of 5 per year). The top left figure shows the amplitudes of the modes at an altitude of 120 km (which can be sounded by optical-imaging systems), while the top right figure shows the amplitudes at an altitude of 150 km (which can be achieved by Doppler sounder). The two bottom figures show the frequencies and quality coefficient of the modes. As no viscosity is taken into account in the model, the Q values for the fundamental and first acoustic modes are high. The Q of the second overtone is much lower because of the partial trapping of the modes, while the Q of the Rayleigh modes is also low due to attenuation in the solid part.

below 120 km does not seem feasible, however. Jupiter normal oscillations do remain coherent for very long times, because of their very high quality coefficient, but on Venus oscillations have a very low Q, typically 200 for periods of 200 s. As a result, the tools used in Jovian seismology, with typical detection threshold in the range of 0.5–1 m s^{-1}, are not sensitive enough to detect atmospheric osciallations on Venus. Amplitudes at higher altitudes of the Venus atmosphere might be detectable with ionospheric sounders, however. The Venus ionospheric structure is indeed thinner, and maximum ionization is found at about 150 km. A sounding from the top is therefore possible at such an altitude, contrary to the sounding situation for Earth, where maximum ionization occurs at about 300–350 km, preventing access from the top of the ionospheric layers between 150 and 200 km altitude. On Venus, electron density ranges from 5×10^3 cm^{-3} to 15×10^3 cm^{-3} during the night and from 5×10^5 cm^{-3} to 5×10^5 cm^{-3} during the day, producing electron plasma frequencies ranging from 0.5–1.5 MHz during the night to 3–7 MHz during the day. A Doppler sounder, with performances below 1 cm s^{-1} in Doppler measurement (Artru et al., 2004), might then offer exciting perspectives for remote sensing seismology on Venus and has to be considered for placement on a future, high-altitude Venus orbiter.

10.03.4.4 Giant Planets Seismology

As is the case with the telluric planets, the internal structure of the giant planets is weakly constrained by geophysical data, and their seismology could provide important new data regarding their structure and dynamics. For a review on the giant planets' internal structures, see Guillot (2005).

We briefly review the main seismological properties of giant planets, before describing the current state-of-the-art techniques for observing these bodies. For a more complete review, including early observations, see Lognonné and Mosser (1993) and Mosser (1995). In contrast to telluric planets, the mantle and possibly even the rocky core of a giant planet are fluid (Guillot, 2005), meaning that only acoustic and gravity waves can be detected. Moreover, the waves are not trapped by a solid-atmospheric discontinuity but rather by the effect of atmospheric density decay, and therefore, the waves are reflected backward only when their frequency is smaller than a frequency called the atmospheric cutoff frequency, which depends on atmospheric structure. Only waves with frequencies lower than the atmospheric cutoff frequency are reflected back into the interior and can therefore generate surface waves or normal modes with high-quality coefficients. In contrast, waves with higher frequencies will continue upward and will be transient, before being dissipated at high altitude.

Figure 28 shows the sound speed and density structure for two typical Jovian interior models, with and without a plasma phase transition (PPT). At this transition, molecular hydrogen is dissociated into metallic or plasma hydrogen (e.g., Saumon and Chabrier, 1989; Saumon et al., 1992). For Jupiter, the cutoff frequency, also shown on **Figure 28**, reaches a maximum of about 4.5 mHz for 50 km above the one-bar level and then decreases to about 3.3 mHz, a stable value achieved between 100 and 200 km of altitude. Low-attenuation normal modes can therefore be found only below 3.3 mHz, while leaky modes can exist between 3.3 and 4.5 mHz. No surface waves exist at frequencies higher than the maximum value of the cutoff frequency; a detailed description of the propagation of seismic waves in the Jovian troposphere and stratosphere can be found in Mosser (1995).

The theoretical spectrum of free Jovian oscillations was computed for models of Jupiter in the late 1970s (Vorontsov and Zharkov, 1981; Vorontsov et al., 1976). As with the Earth and other planets, these theoretical studies, as well as more recent ones (Gudkova and Zharkov, 1997, 1998), including studies of giant exoplanets (Le Bihan and Burrows, 2013), have shown free oscillation frequencies that are highly sensitive to internal structure. As noted earlier, on Jupiter, free oscillations with frequencies higher than about 3 mHz are not trapped. **Figure 29** shows the spectrum of free oscillations for the Jupiter model without PPT, which is shown in **Figure 28**. Only a bulk Q of 10^6 was considered and the viscosity was neglected. As a result, the low Q-values found for frequencies higher than 3 mHz are due mainly to the nontrapping of the waves. Note that the fundamental mode is, in fact, a gravity mode and that the first acoustic mode is for $n=1$ (Lognonné et al., 1994). **Figure 30** shows the amplitudes of the modes, scaled by the square root of the density. Note the trapping of the modes below the cutoff frequency again and the oscillating characters of the acoustic $n=1$ mode for frequencies higher than the cutoff frequency.

The search for the free oscillations of Jupiter started in the late 1980s, with the first successful detection achieved 30 years later (Gaulme et al., 2011). Several techniques were used, including infrared observations (Deming et al., 1989),

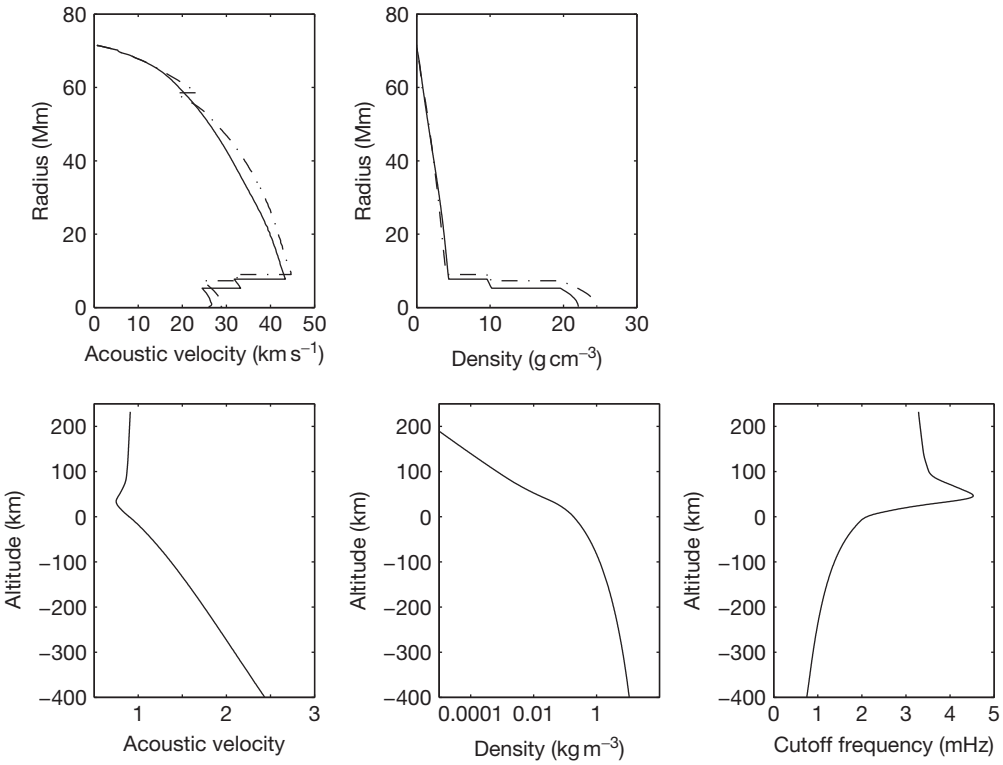


Figure 28 Seismic properties for two different models of Jupiter (Mosser, 1996). The continuous line represents a model without plasma phase transition (PPT) discontinuity, while the dashed line represents a model with a PPT. The two upper figures refer to the entire planet, while the three bottom ones pertain to the atmosphere, displaying the acoustic velocity, the density, and the cutoff frequency, respectively.

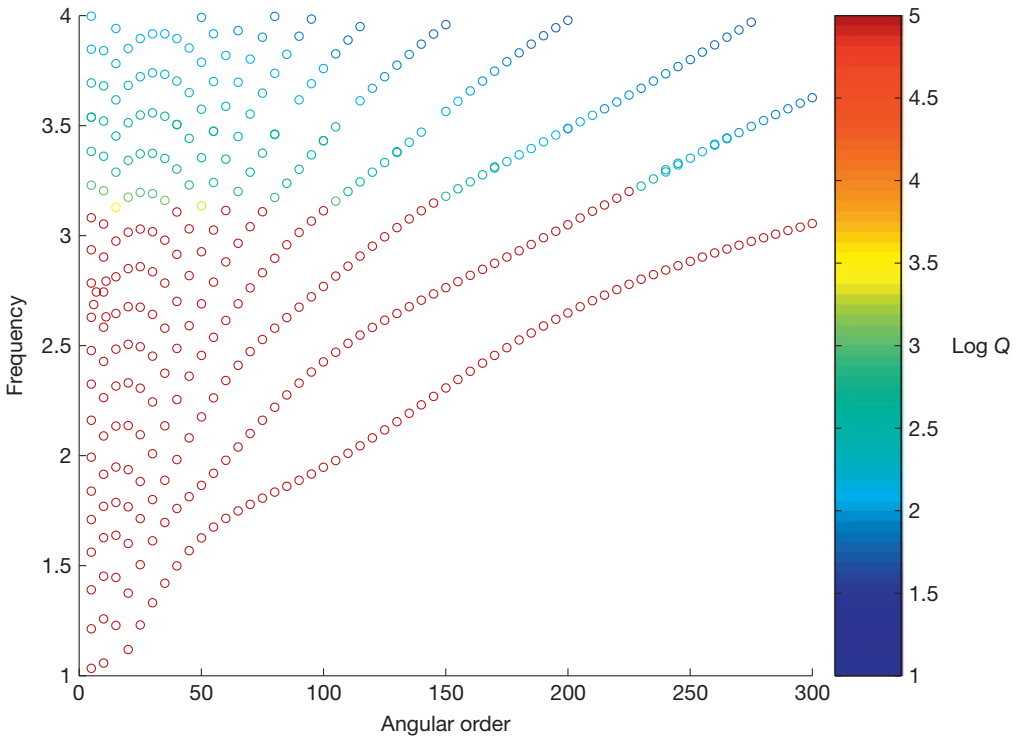


Figure 29 Dispersion curves of the Jovian normal modes for a model of the interior without PPT. Modes are shown with L values increasing by 5 from 5 to 300. The color represents the decimal logarithm of the Q value. Note that Q strongly decreases for frequencies higher than 3 mHz as a consequence of the lack of trapping. The tropospheric guided modes, at about 3.50 mHz, are not shown in the figure.

Doppler spectrometry observations (Cacciani et al., 2001; Mosser et al., 1991; Schmider et al., 1991), Fourier transform interferometry (Mosser et al., 1993, 2000), albedo modulation (Lederer et al., 1995) and Mach–Zender interferometry (Gaulme et al., 2011; Schmider et al., 2003, 2007). This last technique led to the first successful observation of normal mode signatures (Gaulme et al., 2011). Although the latter has not yet resulted in the determination of the frequencies of normal-mode multiplets, several features have been isolated that are associated with low-order modes ($\ell \leq 0-3$). The first is a clear excess of power (over noise) in two detection bands: 0.8–2.1 and 2.4–3.4 mHz. The associated amplitude of normal modes is about 50 cm s^{-1}. This is about twice the detection level of Schmider et al. (1991) and Mosser et al. (1993), who first detected the excess signal over the noise in the frequency band 0.5–1.8 mHz. More importantly, the signal was clearly detected in an Echelon diagram of a periodicity of 155.3 ± 2.2 μHz (An Echelon diagram is the spectrum of an amplitude spectrum and is therefore a tool to detect the relatively regular spacing between normal modes with the same angular order but increasing radial orders). **Figure 30(a)** shows such frequency spacing between modes with increasing radial numbers but the same angular order, suggesting that the detected modes (located in the green box) have radial orders in the range of 5–10.

If these observations do not yet have the precision required for strong constraints on the Jovian interior structure, initial constraints can still be placed on the excitation processes. By using the excitation kernel of normal modes to pressure glut (Lognonné et al., 1994), we can indeed suggest excitations in the 10–10^3-bar level (**Figure 30(b)**), where significant convection is expected to occur. More detailed analysis of these observations will likely be an important part of future Jovian research activities (e.g., Jackiewicz et al., 2012), and we hope that continuous observation will also reduce the detection threshold by an order of magnitude after 1 month, then reaching a limit of about 10 cm s^{-1} and enabling the precise determination of normal-mode frequencies on Jupiter.

As noted earlier, body waves cannot be trapped in the atmosphere, but, in principle, these waves can be excited by a localized source. Due to the large energies involved, the impact of the Shoemaker–Levy 9 comet was such an event. In that case, signal amplitudes were theoretically predicted for different impactor masses (Kanamori, 1993; Lognonné et al., 1994; Marley, 1994). For an impact with an energy $> 10^{21}$ J, peak-to-peak temperature fluctuations greater than ~0.01 K were expected for 10-mHz frequency P waves, while surface waves below 3 mHz were expected to generate fluctuations in excess of 0.01 K for impacts $> 2 \times 10^{21}$ J (Lognonné et al., 1994). No observations were reported by Mosser et al. (1996) for impacts A and H, nor by Walter et al. (1996) for impact R, resulting in upper limits on the impact energy of $1–2 \times 10^{21}$ J. Though the impact did not generate detectable body waves, it did produce a ring-like pattern in the Jovian atmosphere, with two rings propagating at 210 and 450 m s^{-1} (Hammel et al., 1995). If such a wave is neither an acoustic body wave nor a surface wave, it is probably a gravity wave, propagating either in the stratosphere (Walterscheid et al., 2000) or in a deeper layer. In the latter case, an enhancement of the water content by a factor of 10 at a depth of 10 bars (about 80 km below the 1 bar level)

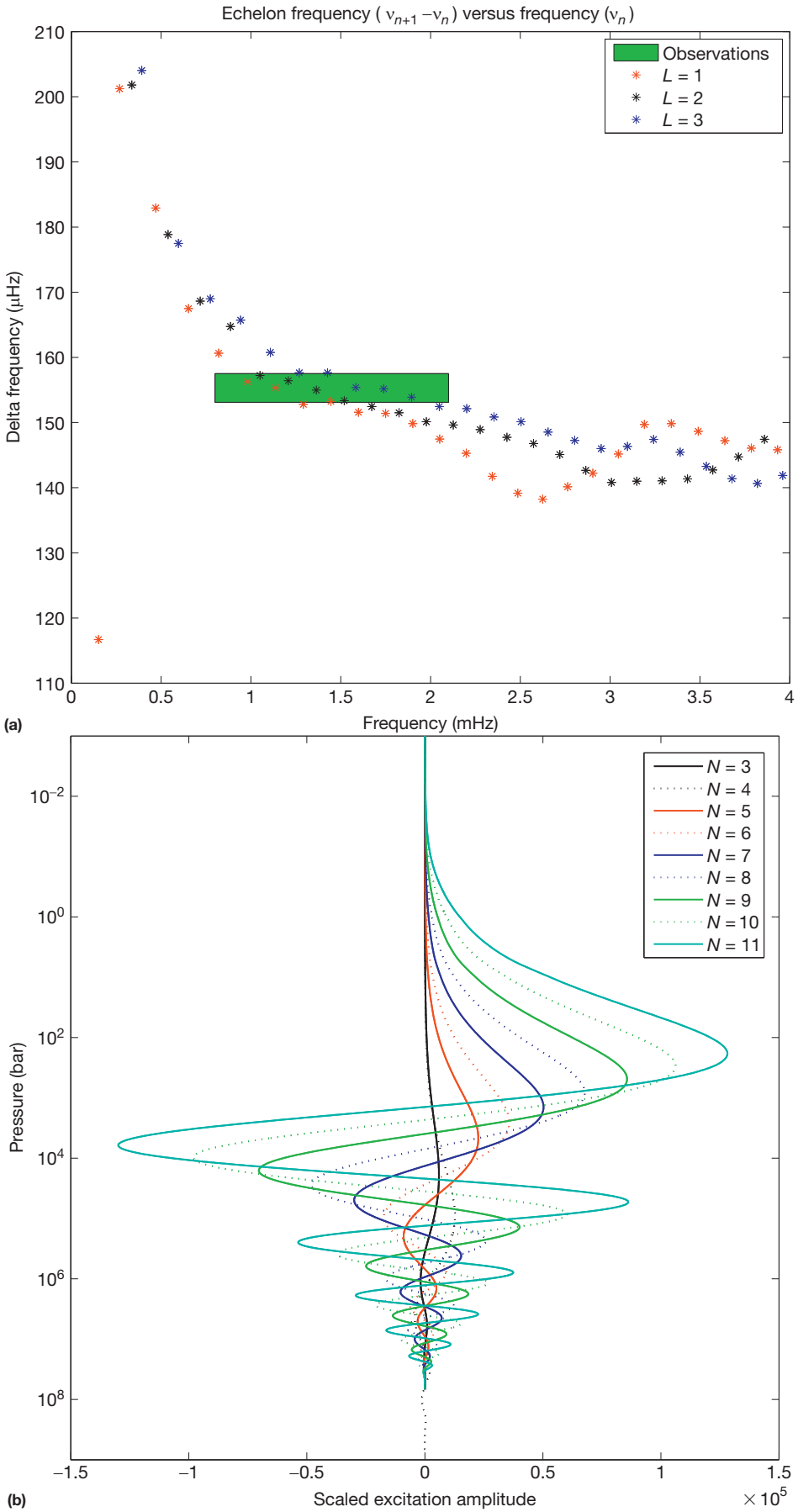


Figure 30 (a) Frequency spacing between two consecutive normal modes for low angular orders $L = 1$, 2, 3, as a function of the normal mode frequency. The green rectangle indicates the periodicity detected by Gaulme et al., 2011. The normal modes start at low frequency with the fundamental mode. (b) Excitation kernel of the $L = 2$ fundamental mode and harmonics, as a function of pressure. The excitation kernel is $\mathbf{u}_{p\_obs}$ $(\mathrm{div}\mathbf{u})_{p\_ex}$, and an observation depth of 1 bar is assumed. These kernels show large amplitudes in the range of 10–1000 bar.

is necessary to explain the wave speed (Ingersoll and Kanamori, 1995; Ingersoll et al., 1994). See Kanamori (2004) for a detailed discussion of this wave.

10.03.5 The New Step: Mars Seismology

We now return to the inner solar system, specifically to martian seismology. In the previous section, we showed that the indirect seismological exploration of Mars may be possible (but very challenging) through the study of the continuous excitation of normal modes by the martian atmosphere. Here we revisit traditional, quake-based seismology, motivated by the InSight Discovery mission, which will deploy a geophysical station on Mars in September 2016 (Banerdt et al., 2012, 2013). InSight will be the first step in almost two decades toward establishing extraterrestrial seismic networks (see, e.g., Lognonné et al., 1996, and for complete list of project references, Lognonné, 2005). A seismic network on Mars would contribute enormously to our understanding of Mars as a planet, and in our opinion, network placement should be one of the highest priorities for the 2020+ exploration of Mars, as illustrated recently by Mocquet et al. (2011) or Dehant et al. (2012).

Given the lack of quake detection by the Viking landers (Anderson et al., 1977a,b) and the failure of the Mars96 mission, which included seismometers on two small autonomous stations and two penetrators, very little is known regarding the seismic properties of Mars. Thus, InSight will provide our next opportunity to directly build our understanding of martian seismology. The mission has been designed with a priori constraints on seismic models that we will review briefly in this section. We then close the section by considering topics relevant to operating a network with a relatively small number (e.g., 2–5) of seismic stations, which is a logical post-InSight next step in Mars' exploration.

10.03.5.1 Interior Structure of Mars

Since the Viking mission, several authors have proposed seismic and density models of Mars. The two main types of constraints used for such models are geophysical and geochemical. Geophysical constraints include the mean density of the planet, which is obtained from the planetary mass (M) and the mean planetary radius (a); the mean moment of inertia factor (I) or the moment of inertia with respect to the rotation axis (C); and the k_2 Love number, which is associated with the gravity signal generated by the Sun tide. Okal and Anderson (1978) proposed a model based on the PREM, with a core size adjusted to fit only the mean density. In this model (and similarly for other models) the pressure range in the entire martian mantle corresponds to the pressure range in Earth's upper mantle.

A second set of geochemical constraints is inferred from the planet's possible bulk composition, which is based on analyses of martian meteorite composition. One of the most-used models, proposed by Dreibus and Wänke (1985), assumes that the bulk silicate composition of Mars is a mixture of carbonaceous C1 chondrite material and volatile-depleted C1 material, with a Mars crustal composition constrained by the SNC group of martian meteorites. Other model mixtures have been proposed, in order to match the $\delta^{17}O/\delta^{18}O$ ratio: a 55% ordinary chondrite H and 45% enstatite chondrite EH (Sanloup et al., 1999), or a 85% H chondrite, 11% CV chondrite, and 4% C1 chondrite (Lodders, 2000; Lodders and Fegley, 1997). One of the important features of these models is the iron enrichment in the mantle (around twice the Earth's value). As illustrated by Mocquet et al. (1996), the presence and depth of seismic discontinuities in the Mars mantle is likely related to the iron content if the mantle is olivine rich (about 60% in volume). A pyroxene-rich mantle (e.g., Sanloup et al., 1999) should display different discontinuities with smaller amplitudes of the seismic jumps (Verhoeven et al., 2005).

Several models have produced revised estimates of the inertia factor, as compared with the factor provided by Sohl and Spohn (1997) and Zharkov and Gudkova (2000), who used a value of $C = 0.365\ \mathrm{Ma}^2$, corrected for the nonhydrostatic contribution of the Tharsis bulge (Kaula, 1979). This value is close to the value later obtained from the measurement of the precession rate of the planet by the Pathfinder mission ($C = (0.3662 \pm 0.0017)\mathrm{Ma}^2$, Folkner et al., 1997).

The new inversions performed during the last decade have found slightly lower value for C, while reducing the error by almost three, using the data collected by the Mars Global Surveyor, MRO, and Odyssey orbiter. $(C/(MR_e^2) = 0.3650 \pm 0.0012$ where R_e is the equatorial radius, Yoder et al., 2003 and $C/(MR_e^2) = 0.3644 \pm 0.0005$ for the most recent determination of Konopliv et al., 2011). All these values should be corrected by the perturbation of the gravitational oblateness J_2 (Munk and MacDonald, 1975) and the one of the J_{22} term (Sohl et al., 2005) in order to get the mean moment of inertia:

$$I = C - \frac{2}{3}J_2\mathrm{Ma}^2 - \frac{8}{3}J_{22}\mathrm{Ma}^2 \qquad [14]$$

These corrections reduce the C value by $1.30e-3$ and $1.68e-4$, respectively (Sohl et al., 2005), leading to $C = (0.3645 \pm 0.0005)\mathrm{Ma}^2$.

The other geophysical datum measured to date is the Love number k_2, obtained from gravity analysis of orbiting spacecraft. This value is poorly known, however. If, following Yoder et al. (2003), Bills et al. (2005), and Konopliv et al. (2011), the value of k_2 is about 0.16 or larger (e.g., 0.164 ± 0.009, Konopliv et al., 2011), then only models with a large core radius (1630–1800 km) are possible (**Figure 31**), as shown by van Hoolst et al. (2003), in contrast to models inferred from the earlier k_2 estimates in the range 0.10–0.13 (Balmino et al., 2005; Marty et al., 2009; Smith et al., 2003). In general, these data support iron cores with some light element(s). In addition to sulfur (e.g., Dreibus and Wänke, 1985), proposed light elements include hydrogen (Gudkova and Zharkov, 2004), silicon (Sanloup et al., 2002; Stevenson, 2001), and carbon (Kuramoto, 1997), while oxygen, another light element possibly present in the Earth's core, is less likely on Mars (Rubie et al., 2004). Differences might, however, be large enough in term of planetary evolution. As an example, the presence of a perovskite-bearing lower mantle becomes impossible for a core larger than 1500 km in radius (e.g., Wang et al., 2013a,b for a recent study), due to insufficient pressures in the mantle, but a larger core radius cannot be ruled out for hot mantle models 4 billion years ago

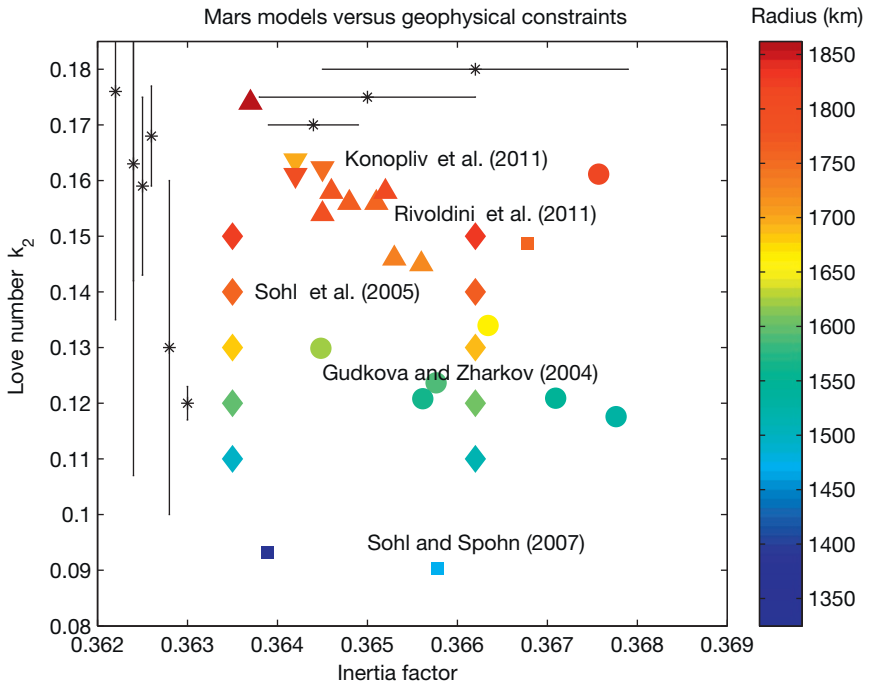


Figure 31 Computed k_2 and inertia factors for the early models of Sohl and Spohn (1997) (squares), Gudkova and Zharkov (2004) (circles), and Sohl et al. (2005) (diamonds) among others, as well as the models produced by Konopliv et al. (2011) (inverted triangle) and Rivoldini et al. (2011) (triangle). This figure illustrates the narrowing zone of the acceptable models resulting from improvements in inertia factor measurement and, to a lesser extent, improvements in k_2 Love number measurement. The inertia factor measurements, located at the top of the figure, are given with their error bar, and from top to bottom, the measurements accredited to Folkner et al. (1997), Yoder et al. (2003), and Konopliv et al. (2011). The measurements for the Love number k_2 are in the left, and they are arranged from left to right, with two types of results displayed. Those measurements in the top converge toward a value of about 0.16 and are reported by Bills et al. (2005), Lemoine et al. (2006), and Konopliv et al. (2011), while those converging toward a smaller value of 0.12–0.13 are reported by Balmino et al. (2005) and Marty et al. (2009). A clear discrepancy exists between these two Love number values as their error bars do not overlap. The core radius is given by the color scale: the upper value of the Love number from Konopliv et al. (2011) suggests a core radius in the range of 1630–1830 km, while the lower value from Marty et al. (2009) suggests a smaller core radius.

(e.g., Rivoldini et al., 2011). This kind of endothermic phase transition from spinel to perovskite has been suggested as a requirement of mantle convective models for the early formation of Tharsis (Harder and Christensen, 1996; van Thienen et al., 2006). The size and state of the core, and the existence and size of a solid inner core, are also critical to understanding the cessation of an early martian dynamo (Acuna et al., 1999, Breuer and Spohn, 2006; Stevenson, 2001). Thus, the determination of the core size is a key objective for seismology, not just for the knowledge of present-day interior structure, but in order to constrain models for the geodynamical evolution of Mars.

The liquid or solid state of the core and mantle attenuation are also parameters that affect seismic-wave propagation and, thus, the design of future seismic experiments on Mars. A liquid (outer) core may result in a seismic shadow zone, as observed on the Earth, and a highly attenuating mantle, due to volatiles and, perhaps, locally warm temperatures, can strongly reduce the amplitudes of the SP body waves, as may be the case on the Moon (Nakamura, 2005). A liquid core is likely and is supported by the Phobos tidal acceleration value (Bills et al., 2005) and by the large k_2 value. As noted by Lognonné and Mosser (1993) and by Zharkov and Gudkova (1997), the low value of the martian Q at the Phobos tidal period implies an unrealistically low intrinsic Q for the planet if the core is solid. In contrast, a Q corresponding to silicate material is found for a model with a liquid core (Lognonné and Mosser, 1993). The extrapolation from the tidal Q to the seismic Q can be performed using either the low frequency dependence Anderson and Given (1982) observed on the Earth, or viscoelastic models (Nimmo and Faul, 2013; Nimmo et al., 2012). In the first case, a shear Q of about 300 at a period of 1 s is found (using a power law of 0.15; Lognonné and Mosser, 1993; Zharkov and Gudkova, 1997), as compared with a Q of 85 at the 5 h 32 min Phobos tide period. On the other hand, the second case yields a lower Q of about 130 at a period of 1 s (Nimmo and Faul, 2013). However, the viscoelastic theory, when used for the Moon, leads to theoretical Q that are up to a factor of two smaller than those seismically observed. Both methods however produce a seismic Q between value of the lower-mantle PREM (Dziewonski and Anderson, 1981) shear Q (about 310) and the upper-mantle shear Q (about 140). As the martian lithosphere is expected to be significantly thicker than Earth's, the intrinsic Q of martian surface waves with periods up to 200 s (which will have most of their energy in the upper 300 km) might be substantially greater than Earth's.

It is unknown whether significant scattering of surface waves will occur on Mars, however.

All other details of mantle structure remain largely unknown. In terms of seismological inferences, a key parameter will be the iron content of the mantle. Mocquet et al. (1996) have studied the effect of increasing iron content from 10% to 40%, and many other models include mantle iron contents ranging from 20% to 30% (Dreibus and Wänke, 1985; Gudkova and Zharkov, 2004; Sanloup et al., 1999; Zharkov and Gudkova, 2005). The main effect of an increase in iron content is to smooth the seismic discontinuities associated with α-olivine-to-β-spinel and β-spinel-to-γ-spinel transitions, as seen in **Figure 32** for both the density and acoustic velocity $v_k(v_k^2 = v_p^2 - 4/3v_s^2)$. For an iron content of 20%, the width of the transition zone at a radius between 2200 and 2400 km is seen to be between 50 and 100 km. Moreover, the discontinuity almost disappears for the high 40% iron content. These transitions contrast with those on Earth, for which both the reduced iron content and the higher pressure gradient result in sharp discontinuities, with a typical width of < 10 km (e.g., Tibi and Wiens, 2005). Reflections from mantle discontinuities on Mars might be less impulsive than reflections on Earth, and as a result, they might only be observable only at LPs, when the seismic wavelength is much greater than the width of the discontinuity, or below frequencies of 0.05–0.1 Hz. Consequently, these phases might be observed only for the largest quakes. In summary, we can expect that, as with the Moon, most of the mantle information will be obtained from direct P and S waves and from phases reflected at the core and the surface (PKP, PcP, ScS, and PP).

10.03.5.2 Martian Seismic Noise

The Viking seismic experiment failed to detect seismic activity largely because of the sensor's wind sensitivity. This wind noise was primarily a consequence of the seismometer's location. The seismometer was deployed on the lander platform and coupled with the ground via the elastic legs of the lander, leading the seismometer to detect lander vibrations due to noise. We can expect increased detection of seismic activity

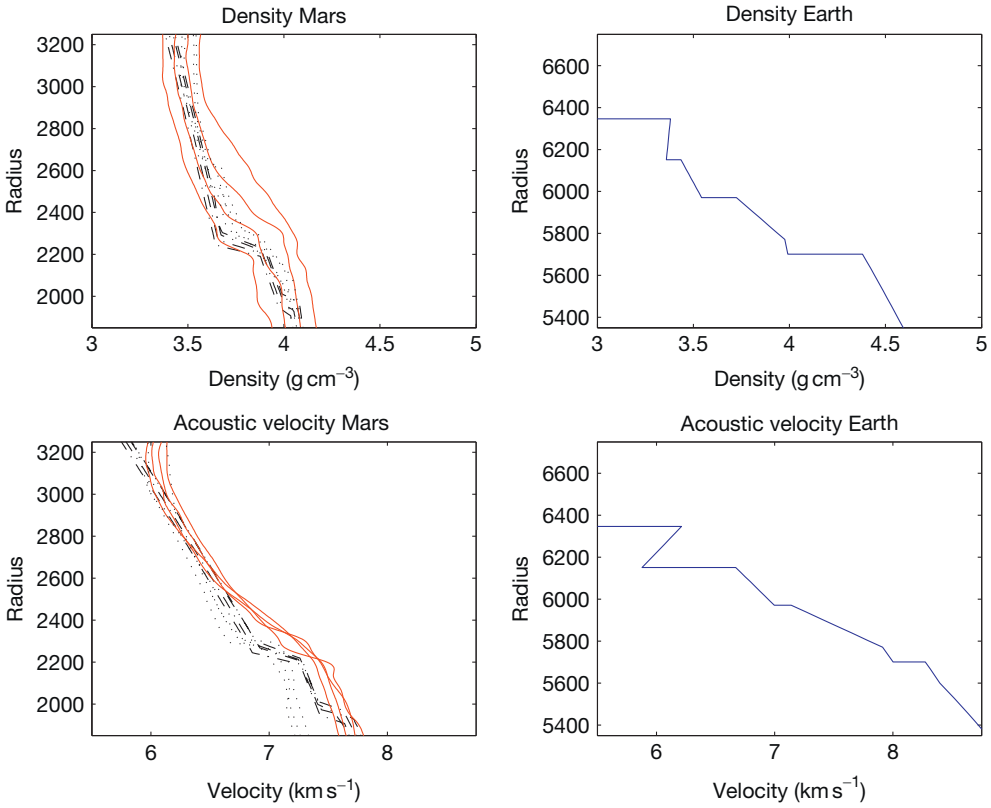


Figure 32 The top-left and bottom-left graphs show the density and acoustic-velocity models M1–M4 of Gudkova and Zharkov (2004), respectively, with Fe contents from 0.20 to 0.25 (0.20–0.22–0.24–0.25) (dashed black lines). In the left graphs, solid red lines represent the models of Mocquet et al. (1996) with Fe contents from 10% to 40% (0.1–0.2–0.3–0.4), and dotted black lines represent the models of Verhoeven et al. (2005) with Fe contents from 0.09 to 0.28 (0.09–0.2–0.25–0.28). An increase in Fe content will smooth the seismic discontinuity and widen the transition zone associated with the α-olivine-to-β-spinel and the β-spinel-to-γ-spinel phase conversions. For comparison, the two right graphs show these parameters for the PREM model (Dziewonski and Anderson, 1981). For the figure on the left for Mars models, the difference in discontinuity depths is mainly related to changes in the core-mantle boundary temperature. A temperature of approximately 2100 K is taken at the core-mantle boundary for the model of Gudkova and Zharkov (2004), whereas a more complex model, with temperature inversion in the mantle and a mantle temperature approximately 500 K colder at the depth of the discontinuities, is taken by Mocquet et al. (1996). These colder temperatures in this model shift the discontinuity to a shallower depth. The temperatures for the models of Verhoeven et al. (2005) are those reported by Breuer and Spohn (2003).

from a future mission, however, given that the seismometer will be directly installed on the ground. Covering the seismometer with a rather simple windshield can strongly reduce wind effects. Field tests have demonstrated the effect of windshielding on Earth at a noise level almost comparable to the low noise model at periods of a few 10s of seconds (Lognonné et al., 1996), and we can therefore expect noise levels as low as 10^{-9} ms^{-2} Hz$^{-1/2}$ in the bandwidth 1–0.05 Hz, which is dominated by microseismic peaks on Earth.

Mars does not have an ocean (the major source of terrestrial noise between 0.07 and 0.14 Hz) or human activity. So, most of that planet's remaining seismic noise will be associated with the interaction of the atmosphere with the subsurface. Lognonné et al. (1996) have shown that, even on Earth, prototype Martian seismometers can reach low noise levels at a quiet site, when protected by a windshield, leaving the overall ground deformation as the ultimate and unescapable noise on a planet with a windy atmosphere. This has been well-documented on Earth, at LPs (Sorrells, 1971; Sorrells et al., 1971) and at SPs of varying durations (Criswell et al., 1975; Steeples et al., 1997; Withers et al., 1996). Wind-generated microseismic noise on Mars has been estimated by Lognonné and Mosser (1993) and later modeled with Large Eddies Simulations by Lognonné et al. (2012), while temperature-generated noise was estimated by van Hoolst et al. (2003).

Ground accelerations produced by the direct deformation of the martian surface due to wind pressure fluctuation could have peak-to-peak amplitudes of a few 10^{-9} m s^{-2}, in the range of 1–0.1 Hz, for wind speeds of ~4 m s^{-1} on non consolidated, regolith-type subsurface (Lognonné and Mosser, 1993; Lognonné et al., 2012). Strictly speaking, these fluctuations are not noise, but the ground deformation generated from the pressure field, and they can therefore be partially corrected by recording the pressure data (e.g., Beauduin et al., 1996).

Temperature variations are the other major source of noise for surface-installed VBB seismometers. However, as shown by van Hoolst et al. (2003), most of the temperature variations on Mars are associated with the daily cycle and can be approximated by a Fourier series with the fundamental mode (of 24 h 56 min) and its harmonics. Moreover, the thermal insulation of space-qualified seismometers can reach time constants larger than 2 h, and an additional wind/thermal shield can create additional attenuation. These two forms of protection reduce the temperature variations experienced by the sensor to $<10^{\circ}$ under martian conditions (Mimoun et al., 2012; Robert et al., 2012), leading to temperature fluctuations of a few tens of milli-Kelvins in the seismic band. An installation with a noise level comparable to terrestrial low noise model (e.g., about 10^{-9} m s^{-2} Hz$^{-1/2}$ in the seismic bandwidth, see Pedersen (1993)) is therefore the realistic goal of InSight.

10.03.5.3 Body-Wave Detection

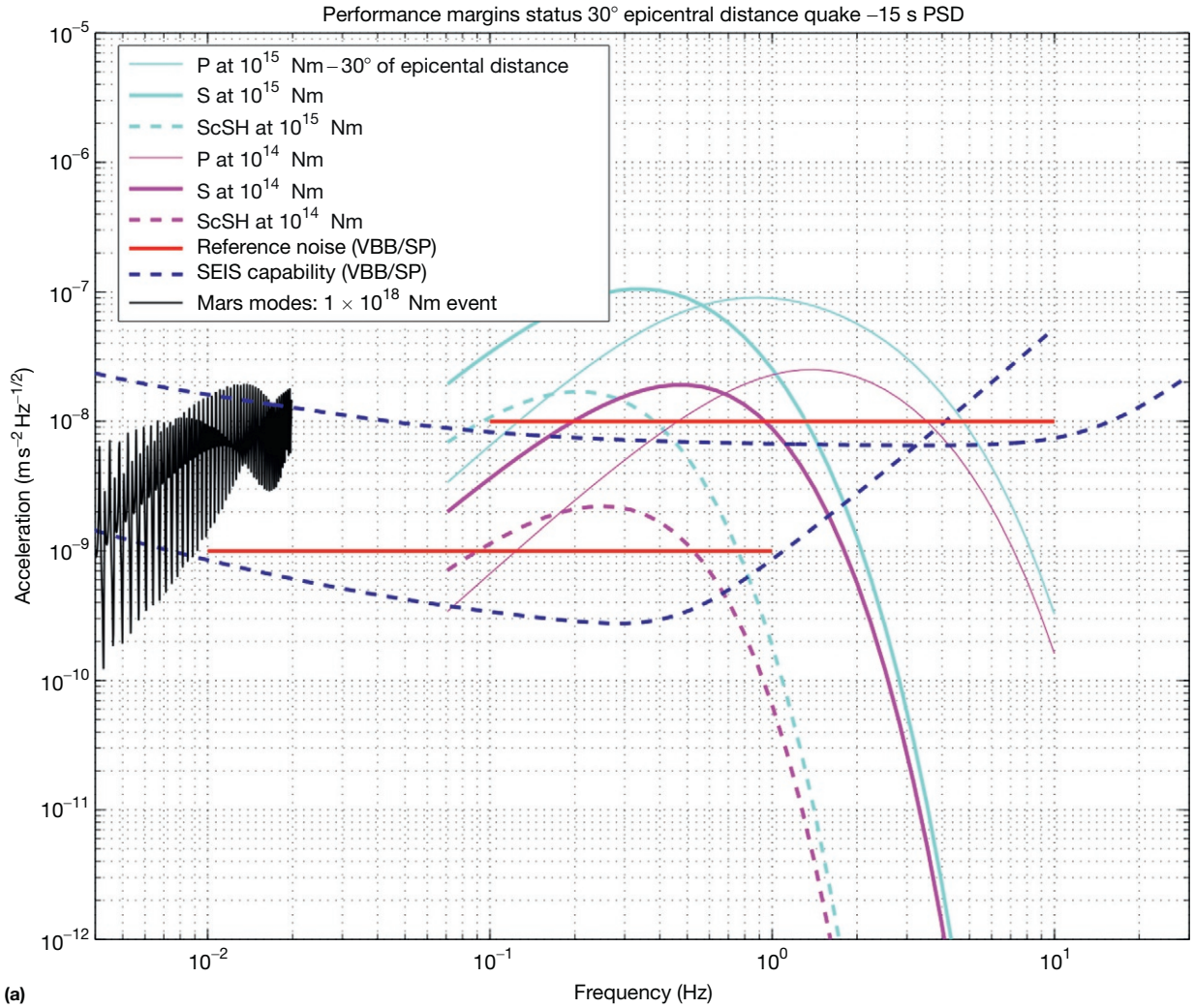
Due to the lower event magnitudes, as compared with event magnitudes on Earth (see **Figure 13** in **Section 10.03.3**), the best signal-to-noise data from Mars will probably be found in the bandwidth of body waves and regional surface waves. Attenuation and the scattering of seismic energy are the main limitations on body-wave detection, and the importance of attenuation on Mars was originally pointed out by Goins and Lazarewicz (1979), who showed that, when operating with a free frequency of 4 Hz, the Viking seismometer was unable to detect remote events due to attenuation. For a Q of 325, which is obtained as explained in **Section 10.03.5.1**, and a velocity of 5 km s^{-1}, S-wave amplitudes are reduced by a factor of 10 after 1200 km of propagation at 1 Hz. Amplitudes at 4 Hz are reduced by the same amount after only 300 km of propagation. The amplitudes of body waves are plotted in more detail in **Figure 33(a)** and **33(b)**, including those of ScSH. The amplitudes are computed for one frequency band (0.3–1.5 Hz) for an isotropic surface shallow source (quake) of seismic moment of 10^{14} and 10^{15} Nm. Transmission across a crust–mantle boundary is taken into account, as are geometrical spreading and attenuation. Sohl and Spohn (1997) model A is used, with both crust and mantle discontinuities, thereby affecting the amplitude for epicentral distance $<20^{\circ}$. While depending on the seismic model used, these amplitudes mainly depend on the attenuation used. Here, the model uses a mantle shear Q provided by $Q_{\mu}=325\times T^{-0.15}$, where T is the period in seconds.

Figure 33(a) shows body-wave amplitudes as a function of frequency at a distance of 30°. Amplitude is relatively independent of epicentral distance only for P waves in the frequency band 0.1–1 Hz. Consequently, this frequency band, with a 1-ng resolution, was chosen for the Optimism seismometer (Lognonné et al., 1998a). At higher frequencies (0.5–2.5 Hz bandwidth), however, amplitudes decrease rapidly with source–receiver distance and are less than S-wave amplitudes at the longest periods. For an instrument with noise characteristics of 10^{-8} ms^{-2} Hz$^{-1/2}$, noise levels of ~2.5 ms^{-2} RMS will be expected in the 0.5–2.5 Hz frequency bandwidth, limiting the detection of S waves at large epicentral distances to quakes with moments larger than 10^{15} Nm.

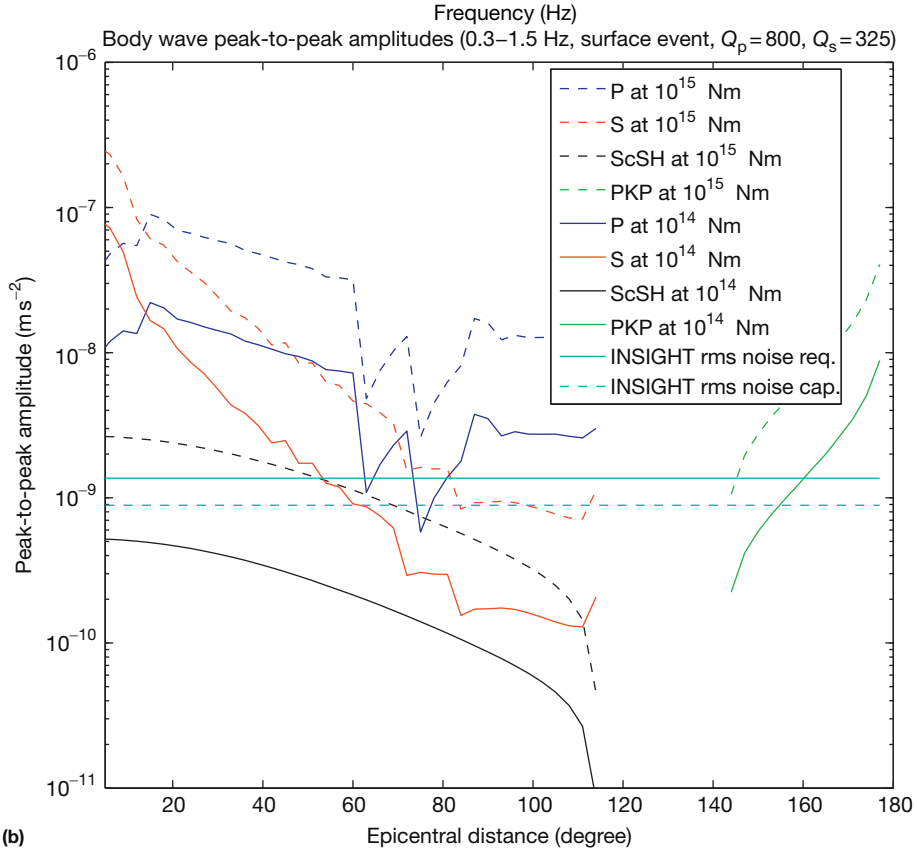
SP body waves are also sensitive to scattering. This relationship has been demonstrated on the Moon and results from intense crustal fracturing due to meteoritic impacts, as well as tectonic structure (e.g., Dainty and Toksöz, 1977; Nakamura, 1977b). On the Earth, scattering is also very strong in volcanic regions. These combined effects suggest that significant scattering may occur in volcanic areas on Mars, particularly in the Tharsis region. For body waves, scattering reduces the peak-to-peak amplitudes by producing conversions (mainly P to SV, P to SH), while spreading this energy in time. This leads to energy transfer from P to S waves up to an equipartition given by $E_p=(v_s^3/v_p^3)(E_s/2)$, where E_p and E_s are the energy of P and S waves, respectively, and v_p and v_s are the velocities of P and S waves (e.g., Aki, 1992; Papanicolaou et al., 1996).

Conversion of P-to-S energy during scattering dominates S-to-P conversion (Aki, 1992). As such, near-surface scattering can greatly reduce the amplitude of P waves both near the source (assuming shallow events) and near the receiver. This amplitude reduction can occur initially as the energy propagates outward from the source into the mantle, and again during the upward propagation from the mantle to the surface, near the receiver. As a result of the shrinking amplitude, the P-wave energy can be reduced by a factor of ten, with larger amplitude reduction due to the length of the coda.

The potential efficiency of a future network configuration can be investigated using these results, as shown by Mocquet


Performance margins status 30° epicentral distance quake –15 s PSD
P at 10^15 Nm – 30° of epicental distance
S at 10^15 Nm
ScSH at 10^15 Nm
P at 10^14 Nm
S at 10^14 Nm
ScSH at 10^14 Nm
Reference noise (VBB/SP)
SEIS capability (VBB/SP)
Mars modes: 1 × 10^18 Nm event
Acceleration (m s^-2 Hz^-1/2)
Frequency (Hz)

(a)


Body wave peak-to-peak amplitudes (0.3–1.5 Hz, surface event, Q_p = 800, Q_s = 325)
P at 10^15 Nm
S at 10^15 Nm
ScSH at 10^15 Nm
PKP at 10^15 Nm
P at 10^14 Nm
S at 10^14 Nm
ScSH at 10^14 Nm
PKP at 10^14 Nm
INSIGHT rms noise req.
INSIGHT rms noise cap.
Peak-to-peak amplitude (m s^-2)
Epicentral distance (degree)

(b)

Figure 33 (Continued)

(1999). Figure 34(a) details such a study for a possible network configuration, assuming total detection for signals with accelerations $>10^{-8}$ m s^{-2} peak-to-peak. They show that a rate of 60% can be achieved for the detection of quakes with seismic moment $>10^{14}$ Nm, which corresponds to Earth moment magnitudes >3.2. As shown in Figure 13, depending on the seismic activity, between 40 and 400 quakes with at least this seismic moment are expected to occur during one martian year. Note that the four-station network configuration, in contrary to the lunar seismic network, will allow a much better characterization of the core, with antipodal stations near Hellas Planitia. Figure 34(b) shows the increase in coverage for a network configuration with an additional station, which would also provide redundancy. Source localization errors will affect the determination of the seismic velocities and might be large for a network with only a few stations. The impact of the number of station used in the location and the source location error has been discussed by Gagnepain-Beyneix et al. (2006), for the lunar case.

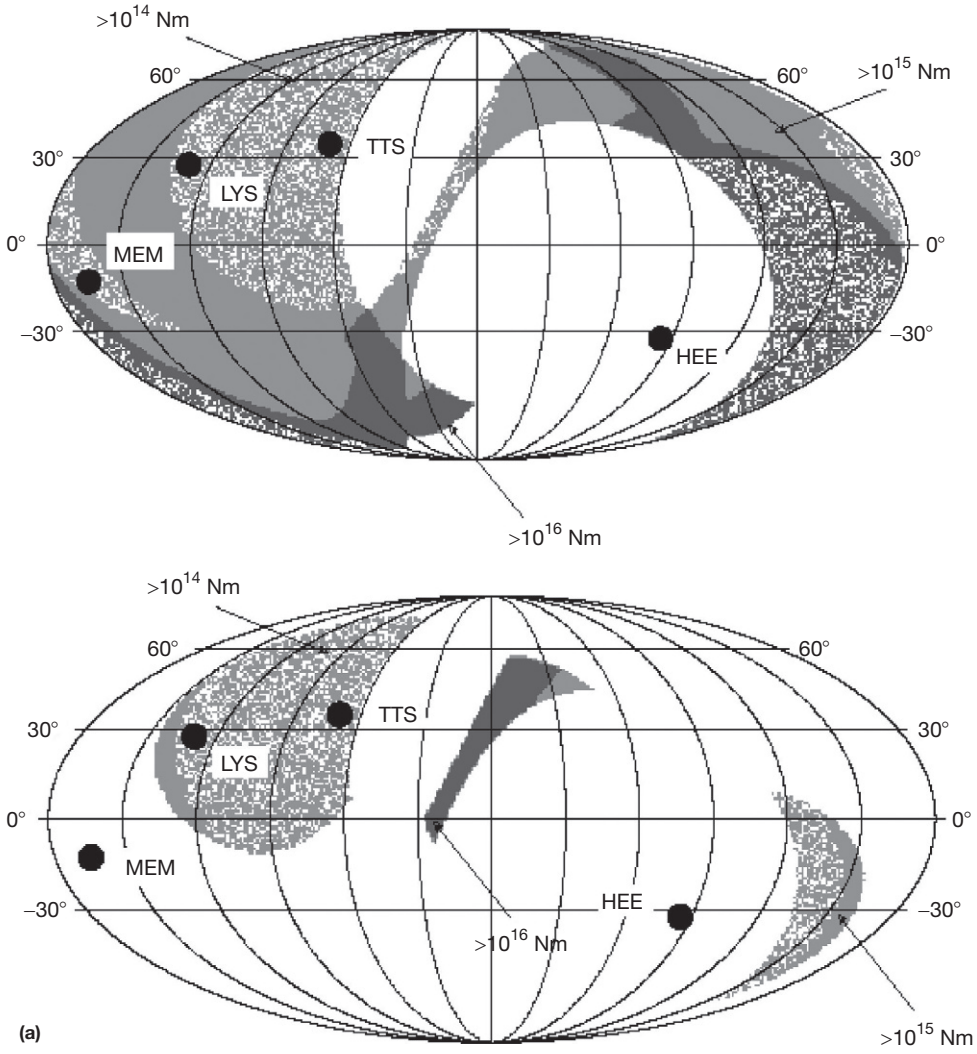


Figure 34 (Continued)

Figure 33 (a) Detection performance of the InSight VBBs and SPs (red line, requirements, black dashed line, capabilities) with respect to the typical amplitude spectral density of ground acceleration generated by marsquakes of 1014-1015 Nm, at 30° of epicentral distance for the model of Sohl and Spohn (1997). Shallow earthquakes might have amplitudes larger by a factor of 2, due to surface amplification. The amplitude power spectrum of P and S body-wave packets have been computed with a duration of 15 s. The amplitudes of normal modes for a 1018 quake are also provided. (b) Peak-to-peak amplitude in the 0.3–1.5 Hz frequency bandwidth with respect to epicentral distance for 10^{14}-Nm (continuous line) and 10^{15}-Nm (long dashed line) seismic moment quakes at the surface. Blue = P, red = S, black = ScSH, and green = PKP. The light blue lines represent the InSight requirement RMS noise and the system's capability in this frequency bandwidth. Large amplitude variations between 60° and 90° are related to the mantle discontinuities of the Sohl and Spohn (1997) model.

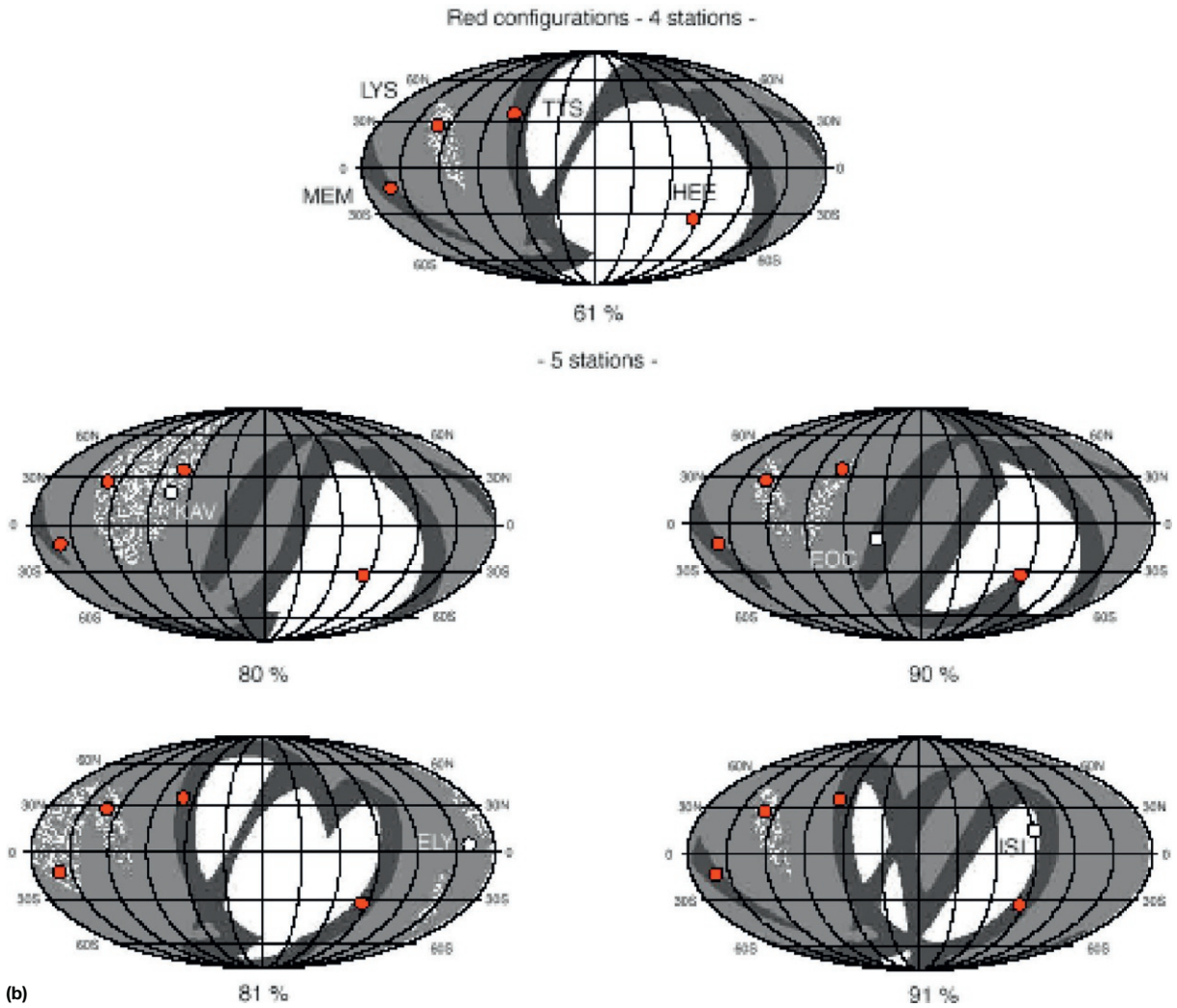


Figure 34 (a) Top: Detection zone of P and S waves for one of the studied configurations in the former Netlander project (MEM, Menomnia Fosae; LYS, Lycus Sulci; TTS, Tempe Terra South; HEE, Hellas East). The white zone to the east of TTS corresponds to the joint shadow zone of MEM and HEE. Note that the detection efficiency is very high in the Tharsis area, where small quakes of seismic moment of 10^{14} Nm might be detected. Bottom: Detection zone of a PKP wave in at least one station and of P and S waves in the three other stations. The success rate is 16%, leading to about 22 quakes detected during the 2 years of operation. Reprinted from Lognonné P, Giardini D, Banerdt B, et al. (2000) The NetLander very broad band seismometer. *Planetary and Space Science* 48: 1289–1302, with permission of Planetary and Space Science. (b) Zone covered by a successful detection of P and S waves on at least three stations for a network with four or five stations. The top figure recalls the detection zones with only four stations, while the other figures represent networks with five stations. In this case, the red dots represent the four nominal landing sites of the Netlander mission. The white dot is a possible new landing site. The shaded zone corresponds (from gray to black) to the detection thresholds from magnitude 3.5–5. Note the large increase of the total surface covered by the network presented below each figure. Note also that the dashed gray zone, corresponding to the detection of small quakes of seismic moment 10^{14} Nm, either has a larger surface or appears in several different spots, each of which is associated with a successive detection on the three closer stations.

10.03.5.4 Normal Mode Excitation and Tidal Observations

With direct body-wave travel-time analysis, the extraction of internal structure information requires precise event localization and, hence, several quakes recorded by at least three stations. For each quake, four data points can then be related to the four source parameters, while the two other points constrain the interior model. Surface-wave and normal-mode analyses have the potential to provide useful information on the interior with fewer stations. As noted above, a cumulative activity of 10^{18}–10^{19} Nm year^{-1} indicates that quakes with moments as large as 10^{18} Nm could be recorded if the station is operated for one martian year or more. The excitation of normal modes by marsquakes was studied by Lognonné et al. (1996) and later by Gudkova and Zharkov (2004). Both studies concluded that normal modes between 5 and 20 mHz with a noise level of 10^{-9} ms^{-2} Hz$^{-1/2}$ could be observed using stacked quake records with a cumulative 10^{18}-Nm moment or from single-record analysis of the greatest quakes (**Figure 35(a)**). Normal modes will peak, with a signal-to-noise ratio of about 10, within Q-cycle windows for such quakes, which correspond to moment magnitudes of 5.5 (Lognonné et al., 1996). Given

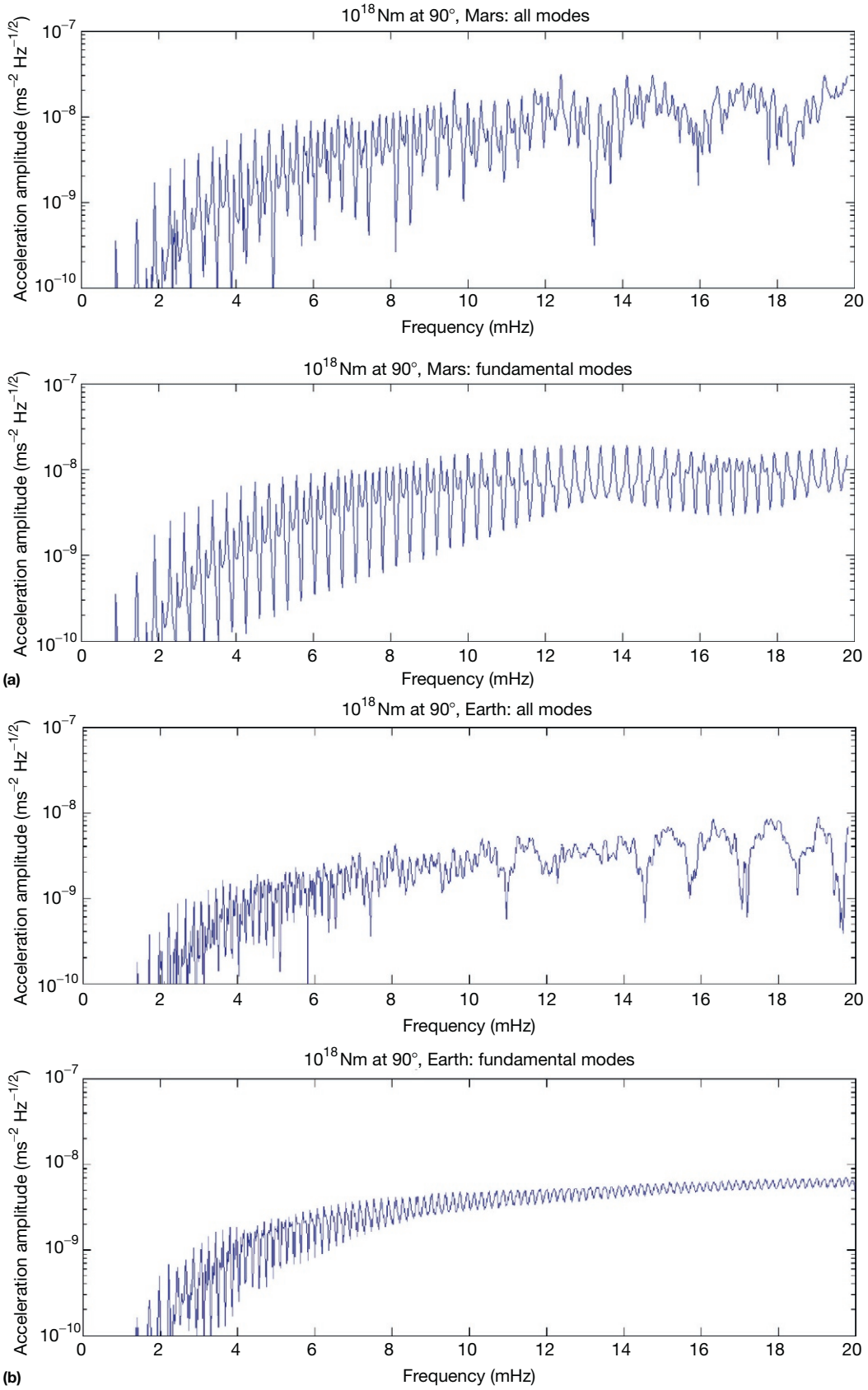

10^{18} Nm at 90°, Mars: all modes
Acceleration amplitude (ms^{-2} Hz$^{-1/2}$)
Frequency (mHz)
10^{18} Nm at 90°, Mars: fundamental modes
(a)
10^{18} Nm at 90°, Earth: all modes
10^{18} Nm at 90°, Earth: fundamental modes
(b)

Figure 35 Acceleration amplitude spectrum for a quake with strike, deep, and rake angles of 45° and a seismic moment of 10^{18} Nm. The azimuth is 20°, and the epicentral distance is 90°. The depth of the event is 30 km. The spectra for Mars are shown in (a). Those for the Earth are in (b). At 5 mHz, the amplitudes exceed 1 ng Hz$^{-1/2}$ and are a factor 4 greater than those observed on the Earth. Note that, below 10 mHz, the fundamental modes constitute most of the signal.

the size of the planet, however, the mode frequencies for a given angular order are typically twice the mode frequencies on Earth (Figure 35(b)). The observation of normal modes below 5 mHz, which are the most sensitive to core size, will be more challenging with such quakes, and modes such as ${}_0S_2$ might remain below the instrument or station noise, especially because of temperature fluctuations due to the operation of the first Mars seismometers at the surface.

As noted by Lognonné and Mosser (1993) and van Hoolst et al. (2003), an alternative core structure constraint will be a surface-sensor measurement of the solid tide amplitude, and, in turn, the associated Love numbers. A successful measurement will be mainly determined by the temperature noise and will require a stack of year-long time series. van Hoolst et al. (2003) have predicted a noise level of about 0.5 K RMS in a frequency bandwidth of 1 mHz around the Phobos tide, based on the Viking temperature data. Both temperature decorrelation and thermal protection can reduce this signal by a factor of 100–250 after data processing, leading to a final noise level of 2.5–1 ngal after one martian year of stacking and a core-determination error equal to 60 km for 1 ngal of noise (Figure 36).

10.03.5.5 Surface Waves

Though the Apollo seismometers recorded no surface waves on the moon, the seismometers bound for Mars are expected to have better LP performance, and the predicted larger magnitude of marsquakes suggests the possibility of surface-wave detections on Mars in the future. By sampling the crust, lithosphere, and upper mantle, surface waves will be an important source of information. Their group velocities, for example, are very sensitive to the crustal thickness, with 10% typical variations for crustal variations of 20 km (Figure 37). The sensitivity of surface waves to the upper mantle will also be important, and the group velocity of surface waves (or the differential group velocity between the fundamental and the overtones) varies by 5–10% for the models of Mocquet et al. (1996) with different iron contents (Figure 32).

Quake localization errors and the effects of lateral variations on the surface waves will likely lead to difficulties in data analysis, however, and these errors will have the greatest impact if only a limited number of stations are deployed, possibly producing errors in the determination of spherically symmetric models of the mantle.

In the case of a single-station approach, as was designed for the InSight project, determining the epicentral distance based on surface waves will be possible after the first detection of the R1, R2, and R3 surface-wave packets generated by a larger quake. The differential travel time between R3 and R1 packets will indeed provide a measure of the averaged propagation velocity of surface waves over the length of the planet's great circle. Once this surface-wave velocity is known, the epicentral distance can be determined as soon as both the R1 and R2 surface-wave packets are defined (Lognonné et al., 2012). This determination will then

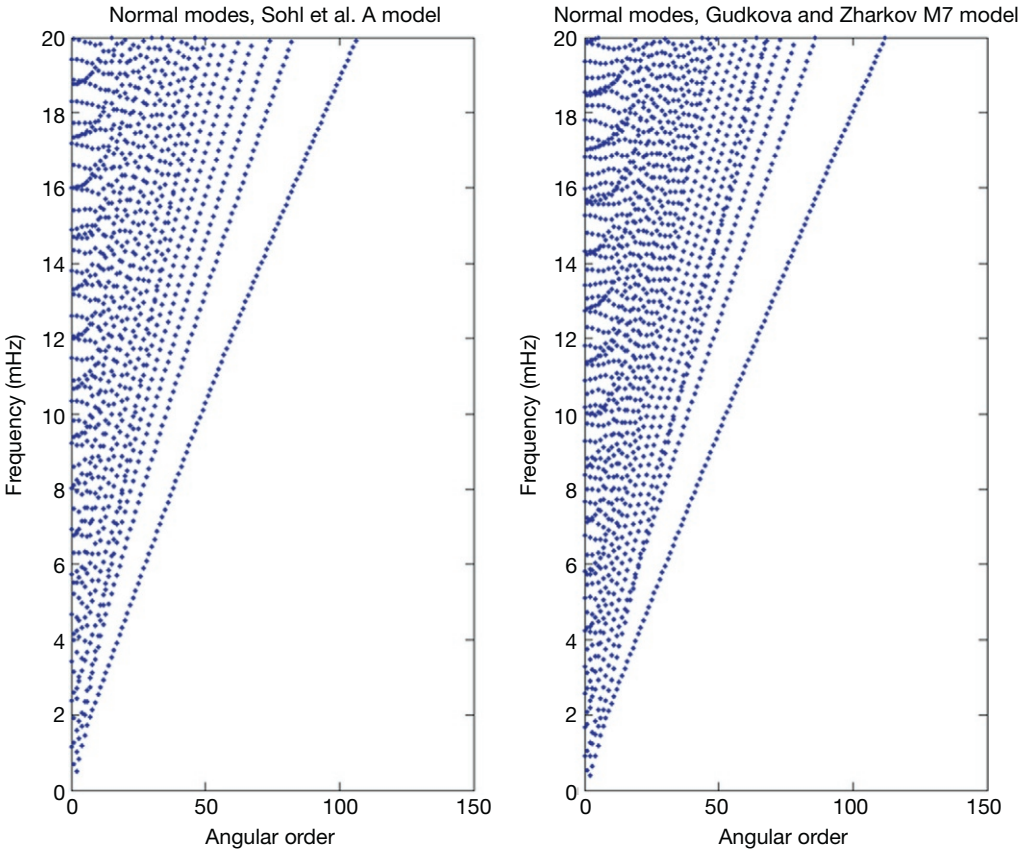


Figure 36 Normal modes for two Mars models: Model A from Sohl and Spohn (1997) and model M7 from Gudkova and Zharkov (2004). Differences in the mode frequencies are about 5–10%.

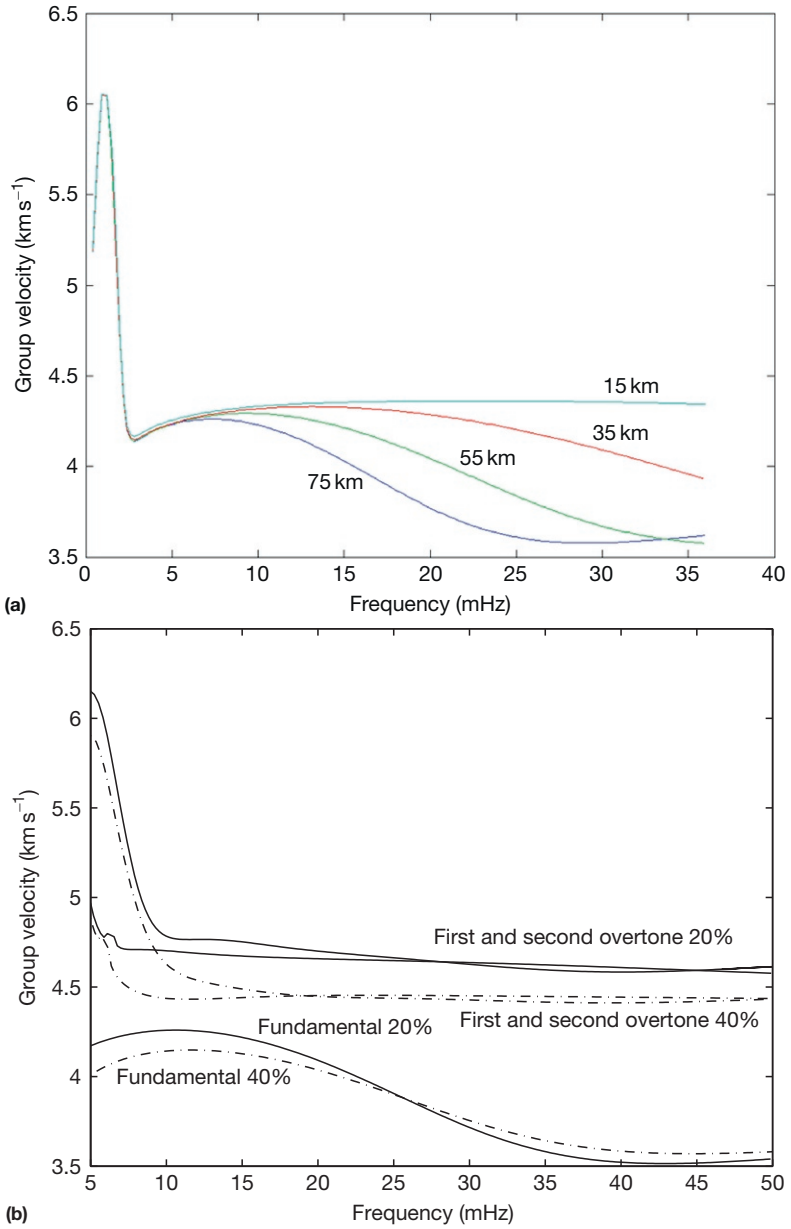


Figure 37 (a) Sensitivity of the Rayleigh waves to the crustal thickness. Model B of Sohl and Spohn (1997) is used, and the crustal velocity and density, which are assumed to be homogeneous, are those of this model. Variations of 10% in the group velocity are found for about 20 km of crustal thickness variation for frequencies in the range of 10–30 mHz. (b) Group velocities for two models of Mocquet et al. (1996) and for the fundamental mode and first and second overtones. The iron content of the mantle is indicated as a percentage. Note that we have added a crust of 50-km thickness to these models, using the Sohl and Spohn (1997) crustal structure. Group velocities without this crust can be found in Lognonné (2005).

enable the creation of interior models through the joint inversion of the R1, R2, P, and S arrival times (Panning et al., 2012).

If two stations are available, the location of quakes via surface waves can be done by measuring the back-azimuth of the surface waves. **Figure 38** shows that, for typical landing sites around the Tharsis bulge, errors in the relative epicentral distance of $<2.5\%$ can be achieved over a wide equatorial band of the planet ($\pm 20^{\circ}$ latitude), when back-azimuth errors of 15° are assumed. Such measurements could resolve the major differences among currently proposed mantle models (**Figure 32**). However, limitations result from the sensitivity of surface-wave azimuth to lateral variation in crustal thickness, and these limitations will prevent the detection of smaller differences, such as those related to the FeO content in the mantle (Mocquet et al., 1996). On Earth, such azimuthal variations typically reach 10°, and it is possible that lateral variations in crustal structure on Mars will lead to larger effects (Larmat et al., 2008). However, as observed on Earth, much of the lateral variations of surface waves can be modeled with a priori constraints on the lithosphere and upper mantle (Nataf and Ricard, 1996). With the recently improved estimates in crustal thickness (Zuber, 2001), lithospheric thickness

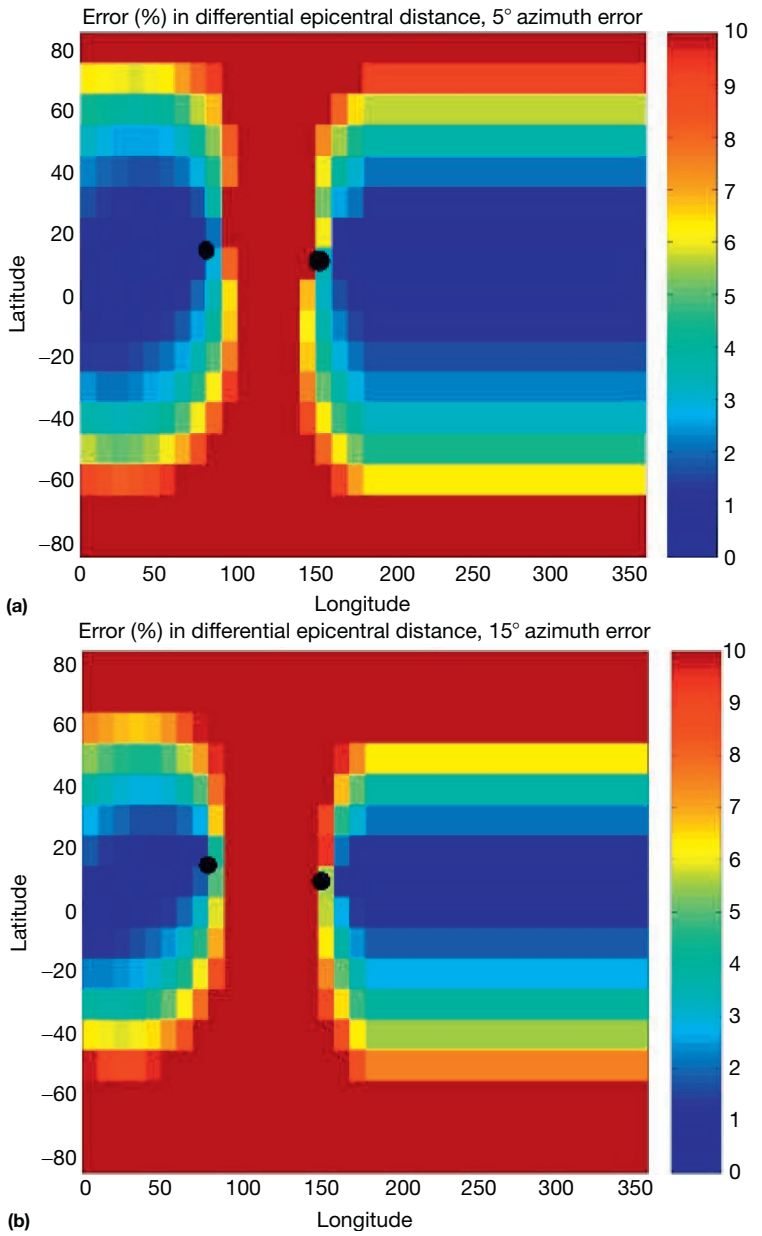


Figure 38 Relative error in the determination of the differential epicentral distance for surface-wave analysis based on the back-azimuth determination in two cases. The differential epicentral distance is the difference between the two epicentral distances (i.e., distance between the epicenter and each station). (a) is for a 5° error, and (b) is for a 15° error. The location of the two stations (filled black circles) is 11° N, 152° W and 15° N, 80° W, and these locations correspond to two landing sites proposed for the NetLander Network, a canceled project. Only locations with an error smaller than 10% are shown, with the latter limit being the a priori difference between the seismic values proposed by the models shown in **Figure 32**.

(Belleguic et al., 2005) and crustal heating, we can expect future developments in modeling the effects of crustal and lithospheric lateral variations on surface waves and associated corrections.

10.03.6 Concluding Remarks

Planetary seismology began almost concurrently with planetary exploration, with seismometers onboard the Ranger probes in the early 1960s. As shown by the Apollo seismic network, a body might be much less seismically active than Earth, but it can still produce several recordable events per day, due to very low noise levels. Both meteorite impacts and moonquakes were recorded on the Moon over a period of more than 7 years. Deep moonquakes remain enigmatic, with repeated brittle failure inferred at specific source locations on the basis of waveform similarity and tidal periodicities. Major questions remain, however. Tidally generated stresses are small relative to those expected to result in failure, and the reasons for brittle failure at depths of 800–1000 km are

not well understood. Large uncertainties regarding the locations of many source regions also exist, and to date, it has not been possible to determine focal mechanisms for these events, although the prominence of shear-wave arrivals in the seismograms suggests that shear failure is occurring.

Meteorites and moonquakes, together with known artificial impacts, have allowed first-order investigations of depth-related variations in seismic velocity. Crustal thickness estimates made shortly after the Apollo missions were suggested to be on the order of 60 km, but more recent studies have revised these estimates to 30–45 km. Seismic-velocity estimates to depths of ~1000 km are possible, but source receiver geometries prohibit direct seismic investigation of lowermost mantle and core structure. Differences among current model estimates for mantle seismic velocities reflect a combination of network geometry, instrument limitations, and likely 3D structure. Thus the seismic exploration of the Moon is far from complete. Large uncertainties remain in the internal structure, and major discontinuities, such as the core–mantle, lower–upper mantle, and mantle–crust transitions, are still not well constrained. Moreover, currently unmapped geographical variations in crustal and mantle structure are a critical key to understanding early lunar evolution.

The work reviewed in this chapter demonstrates that the technology exists to perform ground-based seismic experiments on the Moon and on Mars, and that exciting possibilities also exist for remote detection of seismicity and seismic waves, via atmospheric sounding by satellites. This method could prove fruitful in future studies of Venus, for which long-lived ground-based observations are severely challenging due to surface conditions. In addition, remote detection could allow investigations of the giant planets.

Forty years after Viking, the InSight Discovery mission is expected to reopen the field of planetary seismology by installing a single sensor station on Mars in the fall of 2016. We hope that this station will soon be followed by others, not only on Mars, but on other terrestrial bodies in the solar system. This step is crucial for determining the internal structure of the telluric planets, and it can lead to a better understanding of these bodies and, through comparison, a better understanding of Earth.

Acknowledgments

Catherine Johnson acknowledges the support provided by the Natural Sciences and Engineering Research Council of Canada. Philippe Lognonné acknowledges the support provided by CNES, PNP, and INSU/CNRS, which have been involved with his research for the last 10 years. The authors also thank B. Romanowicz, A. Mocquet, Y. Nakamura, and M. Knapmeyer for their constructive reviews of the original and revised chapters. In addition, the authors thank several colleagues who contributed to this chapter through discussions, figures, and ideas, including B. Banerdt, A. Mocquet, R. Garcia, D.Mimoun, H. Chenet, J. Gagnepain-Beyneix, T. Spohn, M. Knapmeyer, D. Giardini, R. Weber, M. Wieczorek, M. Jellinek, and J.-F. Blanchette-Guertin. A special thanks to all the engineers, technicians, and staff members who have worked to create the instruments used in the various projects described in this paper. This is an IPGP contribution 3608.

References

A'Hearn MF, Belton MJS, Delamere WA, et al. (2005) Deep impact: Excavating comet Tempel 1. *Science* 310: 258–264.

Abercrombie R and Ekström G (2001) Earthquake slip on oceanic transform faults. *Nature* 410: 74–77.

Acuna MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.

Aki K (1992) Scattering conversions P to S versus S to P. *Bulletin of the Seismological Society of America* 82(4): 1969–1972.

Aki K and Richards PG (2002) *Quantitative Seismology*, p. 700. Sausalito, CA: University Science Books.

Anderson DL, Duennebier FK, Latham GV, et al. (1977a) The Viking seismic experiment. *Science* 194(4271): 1318–1321.

Anderson DL and Given J (1982) Absorption band Q model for the Earth. *Journal of Geophysical Research* 87: 3893–3904.

Anderson DL, Miller WF, Latham GV, et al. (1977b) Seismology on Mars. *Journal of Geophysical Research* 82: 4524–4546.

Artru J, Farges T, and Lognonné P (2004) Acoustic waves generated from seismic surface waves: Propagation properties determined from Doppler sounding observation and normal-modes modeling. *Geophysical Journal International* 158: 1067–1077.

Artru J, Lognonné P, and Blanc E (2001) Normal modes modeling of post-seismic ionospheric oscillations. *Geophysical Research Letters* 28: 697–700.

Ball AJ, Lognonné P, Seiferlin K, Spohn T, and Zarnecki JC (2004) Lander and penetrator science for NEO mitigation studies. In: Belton MJS, Morgan TH, and Yeoman DK (eds.) *Mitigation of Hazardous Comets and Asteroid*. Cambridge: Cambridge University Press0521827647, ch. 13.

Balmino G, Duron J, and Marty J (2005) A new martian gravity field mean model and its time variations. American Geophysical Union. Fall Meeting 2005 abstract #G53C.

Banerdt WB, Chui T, Herrin ET, Rosenbaum D, and Teplitz VL (2006) Lunar seismic search for strange quark matter. *Advances in Space Research* 37: 1889–1893.

Banerdt WB, Smrekar S, Alkalai L, et al. (2012) InSight: An integrated exploration of the interior of Mars. In: *43rd Lunar and Planetary Science Conference, The Woodlands, TX, 19–23 March 2012*, LPI Contribution No. 1659, id. 2838B.

Banerdt WB, Smrekar S, Hurst K, et al. (2013) Insight: A discovery mission to explore the interior of mars. In: *44rd Lunar and Planetary Science Conference, The Woodlands, TX, 18–22 March 2013*, LPI Contribution No. 1915.

Beauduin R, Lognonné P, Montagner JP, Cacho S, Karczewski JF, and Morand M (1996) The effect of the atmospheric pressure changes on seismic signals or how to improve the quality of a station. *Bulletin of the Seismological Society of America* 86: 1760–1769.

Beer T (1974) *Atmospheric Waves*, p. 315. New York/London: Halsted Press/Adam Hilger, Ltd.

Belleguic V, Lognonné P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005. http://dx.doi.org/10.1029/2005JE002437.

Benioff H, Press F, and Smith SW (1961) Excitation of the free oscillations of the Earth by earthquake. *Journal of Geophysical Research* 66: 605–620.

Bills BG and Ferrari AJ (1977) A lunar density consistent with topographic, gravitational, librational and seismic data. *Journal of Geophysical Research* 82: 1306–1314.

Bills BG, Neumann GA, Smith DE, and Zuber M (2005) Improved estimate of tidal dissipation within mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004. http://dx.doi.org/10.1029/2004JE002376.

Bills BG and Rubincam DP (1995) Constraints on density models from radial moments: Applications to Earth, Moon, and Mars. *Journal of Geophysical Research* 100: 26305–26316.

Blanchette-Geurtin JF, Johnson CL, and Lawrence JF (2012) Investigation of scattering in lunar seismic coda. *Journal of Geophysical Research: Planets* 117. http://dx.doi.org/10.1029/2011JE004042.

Bois E, Boudin F, and Journet A (1996) Secular variation of the Moon's rotation rate. *Astronomy and Astrophysics* 314: 989–994.

Bouley S, Baratoux D, Vaubaillon J, et al. (2012) Power and duration of impact flashes on the Moon: Implication for the cause of radiation. *Icarus* 218: 115–124.

Bratt SR, Bergman EA, and Solomon SC (1985) Thermoelastic stress: How important as a cause of earthquakes in young oceanic lithosphere. *Journal of Geophysical Research* 90: 10249–10260.

Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics on Mars: Evidence from magnetic field history and crust evolution. *Journal of Geophysical Research: Planets* 108. http://dx.doi.org/10.1029/2002JE001999, CiteID 5072.

Breuer D and Spohn T (2006) Viscosity of the Martian mantle and its initial temperature: Constraints from crust formation history and the evolution of the magnetic field. *Planetary and Space Science* 54: 153–169.
Brown P, Spalding RE, ReVelle DO, Tagliaferri E, and Worden SP (2002) The flux of small near-Earth objects colliding with the Earth. *Nature* 420: 294–296.
Bullen KE (1947) *An Introduction to the Theory of Seismology*. Cambridge: Cambridge University Press.
Bulow R, Johnson CL, and Bills BG (2006) Tidal stress and deep moonquakes. In: *37th Annual Lunar and Planetary Science Conference*, Abstract 1183.
Bulow R, Johnson CL, Bills B, and Shearer P (2007) Temporal and spatial properties of some deep moonquake clusters. *Journal of Geophysical Research* 112: E09003.
Bulow R, Johnson CL, and Shearer P (2005) New events discovered in the lunar Apollo seismic data. *Journal of Geophysical Research: Planets* 110: E10003. http://dx.doi.org/10.1029/2005JE002414.
Burleigh KJ, Melosh HJ, Tornabene LL, Ivanov B, McEwen AS, and Daubar IJ (2012) Impact airblast triggers dust avalanches on Mars. *Icarus* 217: 194–201.
Cacciani A, Dolci M, Moretti PF, et al. (2001) Search for global oscillations on Jupiter with a double-cell sodium magneto-optical filter. *Astronomy and Astrophysics* 372: 317–325.
Cailleau B, Walter TR, Janle P, and Hauber E (2003) Modeling volcanic deformation in a regional stress field: Implications for the formation of graben structures on Alba Patera, Mars. *Journal of Geophysical Research* 108(E12): 5141. http://dx.doi.org/10.1029/2003JE002135.
Calais E and Minster JB (1995) GPS detection of ionospheric perturbations following the January 17, 1994, Northridge earthquake. *Geophysical Research Letters* 22: 1045–1048.
Cameron AG (2000) Higher-resolution simulations of the giant impact. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 133–144. Tucson, AZ: University of Arizona Press.
Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.
Chenet H, Lognonné P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from Apollo seismic dataset. *Earth and Planetary Science Letters* 243: 1–14.
Chuang FC, Beyer RA, McEwen AS, and Thomson BJ (2007) HiRISE observations of slope streaks on Mars. *Geophysical Research Letters* 34: L20204.
Chyba CF, Thomas PJ, and Zahnle KJ (1993) The 1908 Tunguska explosion – Atmospheric disruption of a stony asteroid. *Nature* 361: 40–44.
Cisowski SM, Collinson DW, Runcorn SK, and Stephenson A (1983) A review of lunar paleointensity data and implications for the origin of lunar magnetism. *Journal of Geophysical Research* 88: A691–A704.
Criswell DR, Lindsay JF, and Reasoner DL (1975) Seismic and acoustic emissions of a booming dune. *Journal of Geophysical Research* 80: 4963–4974.
Dainty AM and Toksöz NM (1977) Elastic wave propagation in a highly scattering medium: A diffusion approach. *Journal of Geophysics* 43: 375–388.
Davis PM (1993) Meteoroid impacts as seismic sources on Mars. *Icarus* 105: 469–478.
Davis K and Baker DM (1965) Ionospheric effects observed around the time of the Alaskan earthquake of March 28, 1964. *Journal of Geophysical Research* 70: 1251–1253.
De Pater I and Lissauer JJ (2001) *Planetary Sciences*, p. 66. Cambridge: Cambridge University Press.
Dehant V, Banerdt B, Lognonné P, et al. (2012) Future Mars geophysical observatories for understanding its internal structure, rotation, and evolution. *Planetary and Space Science* 68: 123–145.
Deming D, Mumma MJ, Espenak F, et al. (1989) A search for p-mode oscillations of Jupiter: Serendipitous observations of non-acoustic thermal wave structure. *Astrophysical Journal* 343: 456–467.
Dragoni M and Piombo A (2003) A model for the formation of wrinkle ridges in volcanic plains on Venus. *Physics of the Earth and Planetary Interiors* 135: 161–171.
Dreibus G and Wänke H (1985) A volatile-rich planet. *Meteoritics* 20: 367–382.
Drossart P, Piccioni G, Coradini A, et al. (2004) VIRTIS imaging spectrometer for the ESA/venus express mission, Earth observing systems IX. In: Barnes WL and Butler JJ (eds.). *Proceedings of the SPIE*, vol. 5543, pp. 175–185.
Ducic V, Artru J, and Lognonné P (2003) Ionospheric remote sensing of the Denali Earthquake Rayleigh surface waves. *Geophysical Research Letters* 30: 1951. http://dx.doi.org/10.1029/2003GL017812.
Duennebier F and Sutton GH (1974) Thermal moonquakes. *Journal of Geophysical Research* 79: 4351–4364.
Dziewonski A and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.
Dziewonski AM, Hager BH, and O'Connell RJ (1977) Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research* 82: 239–255.
Edwards WN (2008) Meteor generated infrasound: Theory and observation. In: Le Pichon A (ed.) *Infrasound Monitoring for Atmospheric Studies*, pp. 355–408. New York: Springer-Verlag, chapter 12. http://dx.doi.org/10.1007/978-1-4020-9508-5.
Elkins-Tanton LT, Burgess S, and Yin QZ (2011) The lunar magma ocean: Reconciling the solidification process and geochronology. *Earth and Planetary Science Letters* 304: 326–336.
Ewing M, Latham G, Press F, et al. (1971) Seismology of the Moon and implications on internal structure, origin and evolution. In: De Jager C (ed.) *Highlights of Astronomy*, pp. 155–172. Dordrecht: D. Reidel.
Folkner WM, Yoder CF, Yuan DN, Standisch EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars pathfinder. *Science* 278: 1749–1751.
Forget F, Hourdin F, Fournier R, et al. (1999) Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research* 104: 24155–24176.
Frohlich C and Nakamura Y (2009) The physical mechanisms of deep moonquakes and intermediate-depth earthquakes: How similar and how different? *Physics of the Earth and Planetary Interiors* 173: 365–374.
Fuller M and Stanley MC (1987) Lunar paleomagnetism. *Geomagnetism*, vol. 2, pp. 307–455. New York: Academic Press.
Gagnepain-Beyneix J, Lognonné P, Chenet H, and Spohn T (2006) Seismic models of the Moon and their constraints on the mantle temperature and mineralogy. *Physics of the Earth and Planetary Interiors* 159: 140–166.
Garcia R, Crespon F, Ducic V, and Lognonné P (2005a) 3D ionospheric tomography of post-seismic perturbations produced by the Denali earthquake from GPS data. *Geophysical Journal International* 163: 1049–1064.
Garcia RF, Gagnepain-Beyneix J, Chevrot S, and Lognonné P (2011) Very preliminary reference Moon model. *Physics of the Earth and Planetary Interiors* 188: 96–113.
Garcia R, Lognonné P, and Bonnin X (2005b) Detecting atmospheric perturbations produced by Venus quakes. *Geophysical Research Letters* 32: L16205.
Gaulme P, Schmider FX, Gay J, Guillot T, and Jacob C (2011) Detection of Jovian seismic waves: A new probe of its interior structure. *Astronomy and Astrophysics* 531: A104.
Goins NR (1977) Structure of lunar interior. *Transactions American Geophysical Union* 58(6): 426.
Goins NR, Dainty AM, and Toksöz MN (1981a) Seismic energy release of the Moon. *Journal of Geophysical Research* 86: 378–388.
Goins NR, Dainty AM, and Toksöz MN (1981b) Lunar seismology: The internal structure of the Moon. *Journal of Geophysical Research* 86: 5061–5074.
Goins NR and Lazarewicz AR (1979) Martian seismicity. *Geophysical Research Letters* 6: 368–370.
Golombek MP, Banerdt WB, Tanaka KL, and Tralli DM (1992) A prediction of Mars seismicity from surface faulting. *Science* 258: 979–981.
Gough DO (1986) *Seismology of the Sun and the Distant Stars, Vol. C169: NATO ASI Series. Proceedings of the NATO Advanced Research Workshop*, Cambridge University, England, p. 491. Dordrecht: D Reidel Publishing Co.
Grimm RE and Hess PC (1997) The crust of venus. In: Bougher SW, Hunten DM, Philips RJ, and Tucson AZ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, p. 1205. Tucson, AZ: University of Arizona Press.
Gudkova T, Lognonné P, and Gagnepain-Beyneix J (2011) Seismic source inversion for large impacts detected by the Apollo seismometers. *Icarus* 211: 1049–1065.
Gudkova TV and Zharkov VN (1997) Models of Jupiter and Saturn with water-depleted atmospheres. *Astronomicheskii Vestnik* 31: 99.
Gudkova TV and Zharkov VN (1998) The free oscillations of Jupiter. *Planetary and Space Science* 47: 1211–1224.
Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of the Earth and Planetary Interiors* 142: 1–22.
Guillot T (2005) The interiors of giant planets: Models and outstanding questions. *Annual Review of Earth and Planetary Sciences* 33: 493–530.
Gusev GA and Kurlachev MI (1976) The possibility of observing the lunar microseismic background. *Akademiia Nauk SSSR, Izvestiia, Fizika Zemli*. 84–86 (in Russian).
Hammel HB, Beebe RF, Ingersoll AP, et al. (1995) HST imaging of atmospheric phenomena created by the impact of comet Shoemaker-Levy 9. *Science* 267: 1288–1296.
Hanks TC and Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research* 84: 2348–2350.
Harder H and Christensen UR (1996) A one plume model of Martian mantle convection. *Nature* 380: 507–509.
Hempel S, Knapmeyer M, Jonkers ART, and Oberst J (2012) Uncertainty of Apollo deep moonquake locations and implications for future network designs. *Icarus* 220: 971–980.

Hood LL (1986) Geophysical constraints on the lunar interior. In: Hartmann WK, Philips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 361–410. Houston, TX: Lunar and Planetary Institute.

Hood LL (1995) Frozen fields. *Earth, Moon, and Planets* 67: 131–142.

Hood LL and Jones JH (1987) Geophysical constraints on lunar bulk composition and structure: A reassessment. *Journal of Geophysical Research* 92: E396–E410.

Hood LL, Mitchell DL, Lin RP, Acuna MH, and Binder AB (1999) Initial measurements of the lunar induced magnetic dipole moment using lunar prospector magnetometer data. *Geophysical Research Letters* 26: 2327–2330.

Hood LL and Zuber MT (2000) Recent refinements in geophysical constraints on lunar origin and evolution. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 397–409. Houston, TX: Lunar and Planetary Institute.

Horvath P, Latham GV, Nakamura Y, and Dorman H (1980) Lunar near-surface shear wave velocities at the Apollo landing sites as inferred from spectral amplitude ratios. *Journal of Geophysical Research* 85: 6572–6578.

Hunten DM, Colin L, and Donahue TM (1983) *Venus*, p. 1143. Tucson, AZ: University of Arizona Press.

Ingersoll AP and Kanamori H (1995) Waves from the collisions of comet Shoemaker–Levy-9 with Jupiter. *Nature* 374: 706–708.

Ingersoll AP, Kanamori H, and Dowling TE (1994) Atmospheric gravity waves from the impact of comet Shoemaker–Levy 9 with Jupiter. *Geophysical Research Letters* 21: 1083–1086.

Jackiewicz J, Nettelmann N, Marley M, and Fortney J (2012) Forward and inverse modeling for Jovian seismology. *Icarus* 220: 844–854.

Johnson CL, Stixrude L, Lithgow-Bertelloni C, Bulow RC, and Shearer PM (2005) Mineralogical and seismological models of the lunar mantle. In: *36th Annual Lunar and Planetary Science Conference, League City, TX, 14–18 March 2005*, abstract no. 1565.

Kanamori H (1993) Excitation of Jovian normal modes by an impact source. *Geophysical Research Letters* 20: 2921–2924.

Kanamori H (2004) Some fluid-mechanical problems in geophysics-waves in the atmosphere and fault lubrification. *Fluid Dynamics Research* 34: 1–19.

Kanamori H and Mori J (1992) Harmonic excitation of mantle Rayleigh waves by the 1991 eruption of Mount Pinatubo, Philippines. *Geophysical Research Letters* 19: 721–724.

Kanamori H, Mori J, and Harkrider DG (1994) Excitation of atmospheric oscillations by volcanic eruptions. *Journal of Geophysical Research* 22: 21947–21961.

Kaula WM (1979) The moment of inertia of Mars. *Geophysical Research Letters* 6: 194–196.

Kawamura T, Tanaka S, Kobayashi N, and Lognonné P (2010a) The lunar surface gravimeter as a lunar seismometer. In: *Ground-Based Geophysics on the Moon, Tempe, AZ, 21–22 January 2010*, p. 3023, LPI Contribution No. 1530.

Kawamura T, Kobayashi N, Tanaka S, Lognonné P, and Gagnepain-Beyneix J (2010b) Search for farside deep moonquakes: Source determination of unlocated deep moonquakes with Apollo 17 Lunar Surface Gravimeter. In: *41st Lunar and Planetary Science Conference, The Woodlands, TX, 1–5 March 2010*, p. 1766, LPI Contribution No. 1533.

Kelley M, Livingston R, and McCready M (1985) Large amplitude thermospheric oscillations induced by earthquakes. *Geophysical Research Letters* 12: 577–580.

Kennett BLN and Engdahl ER (1991) Traveltimes for global earthquake location and phase identification. *Geophysical Journal International* 105: 429–465. http://dx.doi.org/10.1111/j.1365-246X.1991.tb06724.x.

Khan A, Connolly JAD, Maclennan J, and Mosegaard K (2007) Joint inversion of seismic and gravity data for lunar composition and thermal state. *Geophysical Journal International* 168: 243–258.

Khan A, Maclennan J, Taylor SR, and Connolly JAD (2006) Are the Earth and the Moon compositionally alike? Inferences on lunar composition and implications for lunar origin and evolution from geophysical modeling. *Journal of Geophysical Research* 111: E05005. http://dx.doi.org/10.1029/2005JE002608.

Khan A and Mosegaard K (2001) New information on the deep lunar interior from an inversion of lunar free oscillation periods. *Geophysical Research Letters* 28: 1791–1794.

Khan A and Mosegaard K (2002) An inquiry into the lunar interior – A non linear inversion of the Apollo seismic data. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JE001658.

Khan A, Mosegaard K, and Rasmussen KL (2000) A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data. *Geophysical Research Letters* 27: 1591–1594.

Khan A, Mosegaard K, Williams JG, and Lognonné P (2004) Does the Moon possess a molten core? Probing the deep lunar interior using results from LLR and Lunar Prospector. *Journal of Geophysical Research* 109: E09007. http://dx.doi.org/10.1029/2004JE002294.

Khavroshkin OB and Tsyplakov VV (1996) Penetrator « Mars-96 », reality and possibilities of seismic experiment, in experimental problems in planetology. 1. United Institute of Physics of the Earth, Russian Academy of Science, Moscow.

Knapmeyer M (2011) Planetary core size: A seismological approach. *Planetary and Space Science* 59: 1062–1068.

Knapmeyer M, Oberts J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars' seismicity. *Journal of Geophysical Research* 111: E11006.

Kobayashi N and Nishida K (1998) Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* 395: 357–360.

Konopliv AS, Asmar SW, Folkner WM, et al. (2011) Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters. *Icarus* 211: 401–428.

Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from Lunar Prospector. *Science* 281: 1476–1480.

Kovach RL and Chyba CF (2001) Seismic detectability of a subsurface ocean on Europa. *Icarus* 150: 279–287.

Kovach RL and Watkins JS (1973a) Apollo 17 seismic profiling-probing the lunar crust. *Science* 180: 1063–1064.

Kovach RL and Watkins JS (1973b) The structure of the lunar crust at the Apollo 17 site. *Geochimica et Cosmochimica Acta* 37(supplement 4): 2549–2560.

Koyama J and Nakamura Y (1980) Focal mechanism of deep moonquakes. In: *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 1855–1865.

Ksanfomaliti LV, Zubkova VM, Morozov NA, and Petrova EV (1982) Microseisms at the Venera 13 and Venera 14 landing sites. *Soviet Astronomical Letters* 8: 241.

Kuramoto K (1997) Accretion, core formation, H and C evolution of the Earth and Mars. *Physics of the Earth and Planetary Interiors* 100: 3–20.

Kuskov OL (1995) Constitution of the Moon: 3. Composition of middle mantle from seismic data. *Physics of the Earth and Planetary Interiors* 102: 239–257.

Kuskov OL (1997) Constitution of the Moon: 4. Composition of the mantle from seismic data. *Physics of the Earth and Planetary Interiors* 90: 55–74.

Kuskov OL and Fabrichanaya OB (1994) Constitution of the Moon: 2. Composition and seismic properties of the lower mantle. *Physics of the Earth and Planetary Interiors* 83: 197–216.

Kuskov OL and Kronrod VA (1998) Constitution of the Moon: 5. Constraints on composition, density, temperature and radius of a core. *Physics of the Earth and Planetary Interiors* 107: 285–306.

Kuskov OL and Kronrod VA (2000) Resemblance and difference between the constitution of the Moon and Io. *Planetary and Space Science* 48: 717–726.

Kuskov OL and Kronrod VA (2009) Geochemical constraints on the model of the composition and thermal conditions of the Moon according to seismic data. *Physics of the Solid Earth* 45: 753.

Kuskov OL, Kronrod VA, and Hood LL (2002) Geochemical constraints on the seismic properties of the lunar mantle. *Physics of the Earth and Planetary Interiors* 134: 175–189.

Lammlein DR (1973) *Lunar Seismicity, Structure and Tectonics*, 99 pp. PhD Dissertation, Columbia University, New York.

Lammlein D (1977) Lunar seismicity and tectonics. *Physics of the Earth and Planetary Interiors* 14: 224–273.

Lammlein D, Latham GV, Dorman J, Nakamura Y, and Ewing M (1974) Lunar seismicity, structure and tectonics. *Reviews of Geophysics and Space Physics* 12: 1–21.

Larmat C, Montagner JP, Capdeville Y, Banerdt WB, Lognonné P, and Vilotte JP (2008) Numerical assessment of the effects of topography and crustal thickness on Martian seismograms using a coupled modal solution–spectral elements method. *Icarus* 196: 78–89. http://dx.doi.org/10.1016/j.icarus.2007.12.030.

Larose E, Khan A, Nakamura Y, and Campillo M (2005) Lunar subsurface investigated from correlation of seismic noise. *Geophysical Research Letters* 32: L16201.

Laster SJ and Press F (1968) A new estimate of lunar seismicity due to meteorite impact. *Physics of the Earth and Planetary Interiors* 1: 151–154.

Latham GV, Ewing M, Dorman J, et al. (1970a) Seismic data from man-made impacts on the Moon. *Science* 170: 620–626.

Latham GV, Ewing M, Dorman J, et al. (1971) Moonquakes. *Science* 174: 687–692.

Latham G, Ewing M, Press F, and Sutton G (1969) The Apollo passive seismic experiment. *Science* 165: 241–250.

Latham GV, Ewing M, Press F, et al. (1970b) Passive seismic experiment. *Science* 167: 455–457.

Le Bihan B and Burrows A (2013) Pulsation frequencies and modes of giant exoplanets. *Astrophysical Journal* 764: 18.

Le Feuvre M and Wieczorek MA (2008) Nonuniform cratering of the terrestrial planets. *Icarus* 197: 291–306.

Lederer SM, Marley MS, Mosser B, Maillard JP, Chanover NJ, and Beebe RF (1995) Albedo features and Jovian seismology. *Icarus* 114: 269–277.
Lee S, Zanolin M, Thode AM, Pappalardo RT, and Makris NC (2003) Probing Europa's interior with natural sound sources. *Icarus* 165: 144–167.
Lehner FE, Witt EO, Miller WF, and Gurney RD (1962) A seismometer for lunar experiments. *Journal of Geophysical Research* 67: 4779–4786.
Leonard RS and Barnes RA Jr. (1965) Observation of ionospheric disturbances following the Alaska earthquake. *Journal of Geophysical Research* 70: 1250.
Lemoine FG, Bruinsma S, Chinn DS, and Forbes JM (2006) Thermospheric studies with Mars Global Surveyor. In: *AIAA/AAS Astrodynamics Specialist Conference, Keystone, CO*. Paper AIAA 2006–6395, 1–13.
Linkin V, Harri AM, Lipatov A, et al. (1998) A sophisticated lander for scientific exploration of Mars: Scientific objectives and implementation of the Mars96 small station. *Planetary and Space Science* 46: 717–737.
Lodders K (2000) An oxygen isotope mixing model for the accretion and composition of rocky planets. *Space Science Reviews* 92: 341–354.
Lodders K and Fegley B (1997) An oxygen isotope model for the composition of Mars. *Icarus* 126: 373–379.
Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 19.1–19.34. http://dx.doi.org/10.1146/annurev.earth.33.092203.122605.
Lognonné P (2009) Seismic waves from atmospheric sources and atmospheric/ionospheric signatures of seismic waves. In: Le Pichon A (ed.) *Infrasound Monitoring for Atmospheric Studies*. New York: Springer-Verlag, chapter 10, http://dx.doi.org/10.1007/978-1-4020-9508-5.
Lognonné P, Artru J, Garcia R, et al. (2006) Ground based GPS tomography of ionospheric post-seismic signal during demeter: The SPECTRE project. *Planetary and Space Science* 54: 528–540.
Lognonné P and Clévédé E (2002) Normal modes of the Earth and Planets. In: Kanamori H, Jennings P, and Lee W (eds.) *International Handbook on Earthquake and Engineering Seismology. International Geophysics Series*, vol. 81A. New York: IASPEI Centennial Publications/Academic Press, chapter 10.
Lognonné P, Clévédé E, and Kanamori H (1998a) Normal mode summation of seismograms and barograms in an spherical Earth with realistic atmosphere. *Geophysical Journal International* 135: 388–406.
Lognonné P, Gagnepain-Beyneix J, Banerdt WB, Cacho S, Karczewski JF, and Morand M (1996) An ultra-broad band seismometer on InterMarsnet. *Planetary and Space Science* 44: 1237–1249.
Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: Implication in terms of structure, formation and evolution. *Earth and Planetary Science Letters* 211: 27–44.
Lognonné P, Giardini D, Banerdt B, et al. (2000) The NetLander very broad band seismometer. *Planetary and Space Science* 48: 1289–1302.
Lognonné P, Le Feuvre M, Johnson CL, and Weber RC (2009) Moon meteoritic seismic hum: Steady state prediction. *Journal of Geophysical Research* 114: E12003.
Lognonné P and Mosser B (1993) Planetary seismology. *Surveys in Geophysics* 14: 239–302.
Lognonné P, Mosser B, and Dahlen FA (1994) Excitation of the Jovian seismic waves by the Shoemaker-Levy 9 cometary impact. *Icarus* 110: 186–195.
Lognonné P, Spiga A, Hurst K, et al. (2012) Martian atmospheric induced micro-seismic noise generation: Large eddy simulations. In: *43rd Lunar and Planetary Science Conference*, The Woodlands, TX, 19–23 March 201. LPI Contribution No. 1659, id. 1994.
Lognonné P, Zharkov VN, Karczewski JF, et al. (1998b) The seismic OPTIMISM experiment. *Planetary and Space Science* 46: 739–747.
Loudin MG and Alexander SS (1978) Observed estimates of spheroidal free oscillations of the Moon and their interpretation. *EOS, Transactions of the American Geophysical Union* 59: 1124.
Mackwell SJ, Zimmerman ME, and Kohlstedt DL (1998) High-temperature deformation of dry diabase with application to tectonics on Venus. *Journal of Geophysical Research* 103: 975–984.
Marley MS (1994) Seimological consequences of the impact of comet Shoemaker-Levy 9. *Astrophysical Journal* 427: L63–L66.
Marty JC, Balmino G, Rosenblatt P, et al. (2009) Martian gravity field model and its time variations from MGS and ODYSSEY data. *Planetary and Space Science* 57: 350–363. http://dx.doi.org/10.1016/j.pss.2009.01.004.
McGarr AG, Latham V, and Gault DE (1969) Meteoroid impacts as sources of seismicity on the Moon. *Journal of Geophysical Research* 74: 5981–5994.
Miller WF (1963) The Caltech digital seismograph. *Journal of Geophysical Research* 68(3): 841–847.
Mimoun D, Wieczorek M, Alkalai L, et al. (2012) Farside explorer: Unique science from a mission to the farside of the Moon. *Experimental Astronomy* 33: 529–585.
Minshull TA and Goulty NR (1988) The influence of tidal stresses on deep moonquake activity. *Physics of the Earth and Planetary Interiors* 53: 41–55.
Mocquet A (1999) A search for the minimum number of stations needed for seismic networking on Mars. *Planetary and Space Science* 47: 397–409.
Mocquet A, Rosenblatt P, Dehant V, and Verhoeven O (2011) The deep interior of Venus, Mars, and the Earth: A brief review and the need for planetary surface-based measurements. *Planetary and Space Science* 59: 1048–1061. http://dx.doi.org/10.1016/j.pss.2010.02.002.
Mocquet A, Vacher P, Grasset O, and Sotin C (1996) Theoretical seismic models of Mars: The importance of the iron content of the mantle. *Planetary and Space Science* 44: 1251–1268.
Mosser B (1995) Propagation and trapping of global oscillations in the Jovian troposphere and stratosphere. *Astronomy and Astrophysics* 293: 586–593.
Mosser B (1996) Giant planets seismology. In: Provost J and Schmider F-X (eds.) *Sounding Solar and Stellar Interiors*, Proceedings of the 181st Symposium of the International Astronomical Union, p. 251. Dordrecht: Kluwer Academic Publishers.
Mosser B, Galdemard P, Lagage P, et al. (1996) Impact seismology: A search primary pressure waves following impacts A and H. *Icarus* 121: 331–340.
Mosser B, Maillard JP, and Mékarnia D (2000) New attempt at detecting the Jovian oscillations. *Icarus* 144: 104–113.
Mosser B, Mekarnia D, Maillard JP, Gay J, Gautierand D, and Delache P (1993) Seismological observations with a Fourier transform spectrometer – Detection of Jovian oscillations. *Astronomy and Astrophysics* 267: 604–622.
Mosser B, Schmider F-X, Delache P, and Gautier D (1991) A tentative identification of Jovian global oscillations. *Astronomy and Astrophysics* 251: 356–364.
Mueller S, Taylor GJ, and Phillips RJ (1988) Lunar composition – A geophysical and petrological synthesis. *Journal of Geophysical Research* 93: 6338–6352.
Munk WH and MacDonald GJF (1975) *The Rotation of the Earth: A Geophysical Discussion*, p. 323. Cambridge: Cambridge University Press.
Nagihara S, Nakamura Y, Lewis LR, et al. (2011) Search and recovery efforts for the ALSEP data tapes. In: *42nd Lunar and Planetary Science Conference*, abstract #1103.
Nakamura Y (1976) Seismic energy transmission in the lunar surface zone determined from signals generated by movement of lunar rovers. *Bulletin of the Seismological Society of America* 66: 593–606.
Nakamura Y (1977a) HFT events: Shallow moonquakes? *Physics of the Earth and Planetary Interiors* 14: 217–223.
Nakamura Y (1977b) Seismic energy transmission in an intensely scattering environment. *Journal of Geophysical Research* 43: 389–399.
Nakamura Y (1978) A1 Moonquakes: Source distribution and mechanisms. In: *Proceedings of the 9th Lunar Science Conference*, pp. 3589–3607.
Nakamura Y (1980) Shallow moonquakes: How they compare with earthquakes. In: *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 1847–1853.
Nakamura Y (1983) Seismic velocity structure of the lunar mantle. *Journal of Geophysical Research* 88: 677–686.
Nakamura Y (2003) New identification of deep moonquakes in the Apollo lunar seismic data. *Physics of the Earth and Planetary Interiors* 139: 197–205.
Nakamura Y (2005) Farside deep moonquakes and deep interior of the Moon. *Journal of Geophysical Research* 110: E01001. http://dx.doi.org/10.1029/2004JE002332.
Nakamura Y (2011) Timing problem with the Lunar Module impact data as recorded by the LSPE and corrected near-surface structure at the Apollo 17 landing site. *Journal of Geophysical Research* 116: E12005.
Nakamura Y and Anderson DL (1979) Martian wind activity detected by a seismometer at Viking lander 2 site. *Geophysical Research Letters* 6: 499–502.
Nakamura Y, Dorman J, Duennebier F, Lammlein D, and Latham G (1975) Shallow lunar structure determined from the passive seismic experiment. *Moon* 13: 57–66.
Nakamura Y, Duennebier FK, Latham GV, and Dorman HJ (1976) Structure of the lunar mantle. *Journal of Geophysical Research* 81: 4818–4824.
Nakamura Y and Koyama J (1982) Seismic Q of the lunar upper mantle. *Journal of Geophysical Research* 87: 4855–4861.
Nakamura Y, Lammlein DR, Latham GV, et al. (1973) New seismic data on the state of the deep lunar interior. *Science* 181: 49–51. http://dx.doi.org/10.1126/science.181.4094.49.
Nakamura Y, Latham GV, and Dorman HJ (1980) How we processed Apollo lunar seismic data. *Physics of the Earth and Planetary Interiors* 21: 218–224.
Nakamura Y, Latham GV, and Dorman HJ (1982) Apollo lunar seismic experiment – Final summary. Proceedings of the 13th Lunar and Planetary Science Conference. *Journal of Geophysical Research* 87: A117–A123.
Nakamura Y, Latham GV, Dorman J, and Harris J (1981) Passive seismic experiment long-period event catalog, final version. Technical Report No. 18. Galveston Geophysics Laboratory, University of Texas at Austin.

Nakamura Y, Latham GV, Dorman HJ, and Harris JE (2008) (revised in 2008 from original in 1981) Passive seismic experiment long-period event catalog. Technical Report 118, revised edition. University of Texas Institute for Geophysics, Austin.

Nakamura Y, Latham GV, Dorman HJ, Ibrahim A-BK, Koyama J, and Horvath P (1979) Shallow moonquakes: Depth, distribution and implications as to the present state of the lunar interior. In: *Proceedings of the 10th Lunar and Planetary Science Conference*, pp. 2299–2309.

Nakamura Y, Dorman HJ, Duennebier FK, Ewing M, Lammlein DR, and Latham GV (1974a) High frequency lunar teleseismic events. In: *Proceedings of the 5th Lunar Science Conference*, pp. 2883–2890.

Nakamura Y, Latham G, Lammlein D, Ewing M, Duennebier F, and Dorman J (1974b) Deep lunar interior inferred from recent seismic data. *Geophysical Research Letters* 1: 137–140.

Nataf H and Ricard Y (1996) 3SMAC: An a priori tomographic model of the upper mantle based on geophysical modeling. *Physics of the Earth and Planetary Interiors* 95: 101–122.

Neal CR, Banerdt WB, Alkalai L, and Lunette Team (2011) Lunette: A two-lander discovery-class geophysics mission to the Moon. In: *42nd Lunar and Planetary Science Conference*, The Woodlands, TX, 7–11 March 2011, p. 2832. LPI Contribution No. 1608.

Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on Mars revealed by the High Resolution Stereo Camera. *Nature* 432: 971–979.

Nimmo F and Faul UH (2013) Dissipation inside Mars at tidal and seismic frequencies. In: *44th Lunar and Planetary Science Conference*, The Woodlands, TX, 18–22 March 2013, p. 2174. LPI Contribution No. 1719.

Nimmo F, Faul UH, and Garnero EJ (2012) Dissipation at tidal and seismic frequencies in a melt-free Moon. *Journal of Geophysical Research* 117: E09005.

Nimmo F and Schenk P (2006) Normal faulting on Europa: Implications for ice shell properties. *Journal of Structural Geology* 28: 2194–2203.

Nishida K, Kobayashi N, and Fukao Y (2000) Resonant oscillations between the solid Earth and the Atmosphere. *Science* 287: 2244–2246.

Oberst J (1987) Unusually high stress drop associated with shallow moonquakes. *Journal of Geophysical Research* 92: 1397–1405.

Okal E and Anderson DJ (1978) Theoretical models for Mars and their seismic properties. *Icarus* 33: 514–528.

Oldham RD (1906) The constitution of the Earth. *Quarterly Journal of Geological Society of London* 62: 456–475.

Ortiz JL, Aceituno FJ, Quesada JA, et al. (2006) Detection of sporadic impact flashes on the Moon: Implications for the luminous efficiency of hypervelocity impacts and derived terrestrial impact rates. *Icarus* 184: 319–326.

Panning M, Lekic V, Manga M, Cammarano F, and Romanowicz B (2006) Long-period seismology on Europa: II. Predicted seismic response. *Journal of Geophysical Research* 111: E12008.

Panning MP, Mocquet A, Beucler E, et al. (2012) InSight: Using Earth data to demonstrate inversion techniques for Mars' interior. In: *43rd Lunar and Planetary Science Conference*, The Woodlands, TX, 19–23 March 2012. LPI Contribution No. 1659, id. 1515.

Papanicolaou G, Ryzhik L, and Keller JB (1996) Stability of the P to S-energy ratio in the diffusive regime. *Bulletin of the Seismological Society of America* 86: 1107–1115.

Pedersen J (1993) Observation and modeling of seismic background noise. *U.S. Geological Survey Open-File Report*, 93-322, pp. 1–45.

Phillips RJ (1991) Expected rate of marsquakes in scientific rationale and requirements for a global seismic network on mars, pp. 35–38. LPI Technical Report, 91-02. Houston, TX: Lunar and Planetary Institute.

Poveda A, Herrera MA, García JL, Hernández-Alcántara A, and Curioca K (1999) The expected frequency of collisions of smallmeteorites with cars and aircraft. *Planetary and Space Science* 47: 715–719.

Press F, Buwalda P, and Neugebauer M (1960) A lunar seismic experiment. *Journal of Geophysical Research* 65: 3097–3105.

Qin C, Muirhead AC, and Zhong S (2012) Correlation of deep moonquakes and mare basalts: Implications for lunar mantle structure and evolution. *Icarus* 220: 100–105.

ReVelle D (1976) On meteor-generated infrasound. *Journal of Geophysical Research* 81: 1217–1229.

Rhie A and Romanowicz B (2004) Excitations of the Earth's incessant free oscillation by atmosphere/ocean/solid Earth coupling. *Nature* 431: 552–556.

Rhie J and Romanowicz B (2006) A study of the relation between ocean storms and the Earth's hum. *Geochemistry Geophysics Geosystems* 7(10): Q10004.

Richardson JE, Melosh HJ, and Greenberg R (2004) Impact-induced seismic activity on asteroid 433 Eros: A surface modification. *Science* 306: 1526–1529.

Riedel KS and Sidorenko A (1995) Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing* 43: 188–195.

Righter K (2002) Does the moon have a metallic core? Constraints from giant impact modeling and siderophile elements. *Icarus* 158: 1–13.

Rivoldini A, Van Hoolst T, Verhoeven O, Mocquet A, and Dehant V (2011) Geodesy constraints on the interior structure and composition of Mars. *Icarus* 213: 451–472. http://dx.doi.org/10.1016/j.icarus.2011.03.024.

Roark S, Cottingham C, Dissly R, Scheeres D, Petr V, and Housen K (2010) Explosive surface pods for cratering experiments on small bodies. In: *41st Lunar and Planetary Science Conference*, Woodlands, TX, 1–5 March 2010, p. 2100. LPI Contribution No. 1533.

Robert O, Gagnepain-Beyneix J, Nebut T, et al. (2012) The InSight Very Broad Band (VBB) seismometer payload. In: *43rd Lunar and Planetary Science Conference*, The Woodlands, TX, 19–23 March 2012. LPI Contribution No. 1445, id. 2025.

Rubie DC, Gessmann CK, and Frost DJ (2004) Partitioning of oxygen during core formation on the Earth and Mars. *Nature* 429: 58–61.

Sanloup C, Guyot F, Gillet P, and Fei Y (2002) Physical properties of liquid Fe alloys at high pressure and their bearings on the nature of metallic planetary cores. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000808, ECV4-1-4-9.

Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of the Earth and Planetary Interiors* 112: 43–54.

Saumon D and Chabrier G (1989) Fluid hydrogen at high density: The plasma phase transition. *Physical Review Letters* 62: 2397–2400.

Saumon D, Hubbard WB, Chabrier G, and van Horn HM (1992) The role of the molecular-metallic transition of hydrogen in the evolution of Jupiter, Saturn, and brown dwarfs. *Astrophysical Journal* 391: 827–831.

Schmider FX, Gay J, Gaulme P, et al. (2007) SYMPA, a dedicated instrument for Jovian seismology – I. Principle and performance. *Astronomy and Astrophysics* 474: 1073–1080.

Schmider FX, Gay J, Jacob C, et al. (2003) SYMPA: A dedicated instrument and a network for seismology of giant planets. *Astrophysics and Space Science* 284: 449–452.

Schmider FX, Mosser B, and Fossat E (1991) A possible detection of Jovian global oscillations. *Astronomy and Astrophysics* 248: 281–291.

Scholz CH (1990) *The Mechanism of Earthquakes and Faulting*. Cambridge: Cambridge University Press.

Schultz PH, Hermalyn B, and Veverka J (2013) The deep impact crater on 9P/Tempel-1 from Stardust-NExT. *Icarus* 222: 502–515.

Seidensticker KJ, Möhlmann D, Apaty I, et al. (2007) SESAME – An experiment of the Rosetta lander Philae: Objectives and general design. *Space Science Reviews* 128: 301–337.

Sens-Schönfelder C and Larose E (2010) Lunar noise correlation, imaging and monitoring. *Earthquake Science* 23: 519–530.

Shapiro NM and Campillo M (2004) Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters* 31: L07614. http://dx.doi.org/10.1029/2004GL019491.

Shapiro NM, Campillo M, Stehly L, and Ritzwoller MH (2005) High-resolution surface-wave tomography from ambient seismic noise. *Science* 307: 1615–1618.

Shearer P (1999) *Introduction to Seismology*, p. 272. Cambridge: Cambridge University Press.

Shoemaker EM, Wolfe RF, and Shoemaker CS (1990) Asteroid and comet flux in the neighborhood of earth. Geological Society of America, Special Paper 247, 155–170, http://dx.doi.org/10.1130/SPE247-p155.

Smith DE, Zuber MT, Torrence MH, and Dunn PJ (2003) Estimating the k2 tidal gravity love number of Mars. American Geophysical Union, Fall Meeting 2003, abstract #P31A-05.

Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2005JE002520.

Sohl F and Spohn T (1997) The interior structure of Mars: Implications from SNC meteorids. *Journal of Geophysical Research* 102: 1613–1636.

Solomon S, Bullock M, and Grinspoon D (1999) Climate change as a regulator of tectonics on Venus. *Science* 286: 87–90.

Sorrells GG (1971) A preliminary investigation into the relationship between long-period seismic noise and local fluctuations in the atmospheric pressure field. *Geophysical Journal of the Royal Astronomical Society* 26: 71–82.

Sorrells GG, McDonald JA, Der ZA, and Henin E (1971) Earth motion caused by local atmospheric pressure changes. *Geophysical Journal of the Royal Astronomical Society* 26: 83–98.

Spohn T, Acuña MH, Breuer D, et al. (2001a) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.

Spohn T, Konrad W, Breuer D, and Ziethe R (2001b) The longevity of lunar volcanism: Implications of thermal evolution calculations with 2D and 3D mantle convection models. *Icarus* 149: 54–65.

Steeples DW, Schmeissner C, and Macy B (1997) Recording wind microstructure with a seismograph. *Geophysical Research Letters* 24: 2375–2377.
Stevenson DJ (2001) Mars' core and magnetism. *Nature* 412: 214–219.
Stofan ER, Saunders RS, Senske D, et al. (1993) Venus interior structure mission (VISM): Establishing a seismic network on Venus. In: *Workshop on Advanced Technologies for Planetary Instruments*, Part 1, pp. 23–24, SEE N93-28764 11–91. Houston, TX: Lunar and Planetary Institute.
Suda N, Nawa K, and Fukao Y (1998) Incessant excitation of the Earth's free oscillations. *Science* 279: 2089–2091.
Suggs RM, Cooke WJ, Suggs RJ, Swift WR, and Hollon N (2008) The NASA lunar impact monitoring program. *Earth, Moon, Planets* 102: 293–298.
Surkov Y (1990) *Exploration of Terrestrial Planets from Spacecraft*, vol. 390, pp. 372–373. New York: Ellis, Horwood.
Sutton GH and Latham GV (1964) Analysis of a feedback controlled seismometer. *Journal of Geophysical Research* 69: 3865–3882.
Sutton GH and Steinbacher R (1967) Surveyor seismograph experiment. *Journal of Geophysical Research* 72: 841–844.
Tanaka S, Mitani T, Otake H, et al. (2013) Present status of the Lunar Lander Project SELENE-2. In: *44th Lunar and Planetary Science Conference*, The Woodlands, TX, 18–22 March 2013, p. 1838. LPI Contribution No. 1719.
Tanimoto T (1999) Excitation of normal modes by atmospheric turbulence: Source of long period seismic noise. *Geophysical Journal International* 136: 395–402.
Tanimoto T (2001) Continuous free oscillations: Atmosphere-solid Earth coupling. *Annual Review of Earth and Planetary Sciences* 29: 563–584.
Tanimoto T (2005) The oceanic excitation hypothesis for the continuous oscillations of the Earth. *Geophysical Journal International* 160: 276–288. http://dx.doi.org/10.1111/j.1365-246X.2004.02484.x.
Tanimoto T, Eitzel M, and Yano T (2008) The noise cross-correlation approach for Apollo 17 LSPE data: Diurnal change in seismic parameters in shallow lunar crust. *Journal of Geophysical Research* 113: E08011.
Tanimoto T, Um J, Nishida K, and Kabayashi N (1998) Earth's continuous oscillations observed on seismically quiet days. *Geophysical Research Letters* 25(10): 1553–1556.
Taylor SR (1982) *Planetary Science: A Lunar Perspective*, p. 502. Houston, TX: Lunar and Planetary Institute.
Teenby NA and Wookey J (2011) Seismic detection of meteorite impacts on Mars. *Physics of the Earth and Planetary Interiors* 186: 70–80.
Tibi R and Wiens DA (2005) Detailed structure and sharpness of upper mantle discontinuities in the Tonga subduction zone from regional broadband arrays. *Journal of Geophysical Resarch* 110(B6).
Tobias RL (1978) *The Lunar Surface Gravimeter and the Search for Gravitational Radiation*. PhD Thesis, Maryland University, College Park.
Toksöz MN (1974) Geophysical data and the interior of the Moon. *Annual Review of Earth and Planetary Sciences* 2: 151.
Toksöz MN, Goins NR, and Cheng CH (1977) Moonquakes: Mechanisms and relations to tidal stresses. *Science* 196: 979–981.
Tolstoy M, Vernon FL, Orcutt JA, and Wyatt FK (2002) The breathing of the seafloor: Tidal correlations of seismicity on axial volcano. *Geology* 30: 503–506.
van Hoolst T, Dehant V, Roosbeek F, and Lognonné P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 161: 281–296.
van Thienen P, Rivoldini A, van Hoolst T, and Lognonné P (2006) A top-down origin for Martian mantle plumes. *Icarus* 185: 197–201.
Verhoeven O, Rivoldini A, Vacher P, et al. (2005) Interior structure of terrestrial planets: Modeling Mars' mantle and its electromagnetic, geodetic, and seismic properties. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2004JE002271.
Vinnik L, Chenet H, Gagnepain-Beyneix J, and Lognonné P (2001) First seismic receiver functions on the Moon. *Geophysical Research Letters* 28: 3031–3034.
Von Rebeur-Paschwitz E (1889) The earthquake of Tokyo, April 18, 1889. *Nature* 40: 294–295.
Vorontsov SV and Zharkov VN (1981) Natural oscillations of the giant planets: Influence of rotation and ellipticity. *Astronomicheskii Zhurnal* 58: 1101–1114 (also in *Sov. Astron.*, 25, 627–634, 1982).
Vorontsov SV, Zharkov VN, and Lubimov VM (1976) The free oscillations of Jupiter and Saturn. *Icarus* 27: 109–118.
Waenke H, Baddenhausen H, Blum K, et al. (1977) On the chemistry of lunar samples and achondrites – Primary matter in the lunar highlands: A re-evaluation. In: *Proceedings of the 8th Lunar Science Conference*, vol. 2, pp. 2191–2213.
Walker JD and Huebner WF (2004) Seismological investigation of asteroid and comet interiors. In: Belton MJS, Yeomans DK, and Morgan TH (eds.) *Mitigation of Hazardous Impacts Due to Asteroids and Comets*, pp. 234–265. Cambridge: Cambridge University Press0521827647.
Walter CM, Marley MS, Hunten DM, et al. (1996) A search for seismic waves from the impact of the SL/9 R fragment. *Icarus* 121: 341–350.
Walterscheid RL, Brinkman DG, and Schubert G (2000) Wave disturbances from the comet SL-9 impacts into Jupiter's atmosphere. *Icarus* 145: 140–146.
Wang Y, Takenaka H, Jiang X, and Lei J (2013a) Modelling two-dimensional global seismic wave propagation in a laterally heterogeneous whole-Moon model. *Geophysical Journal International* 192: 1271–1287.
Wang Y, Wen L, and Weidner DJ (2013b) Composition of Mars constrained using geophysical observations and mineral physics modeling. *Physics of the Earth and Planetary Interiors* 224: 68–76. http://dx.doi.org/10.1016/j.pepi.2013.08.005.
Ward SN (1998) On the consistency of earthquake moment rates, geological fault data, and space geodetic strain: The United States. *Geophysical Journal International* 134: 172–186.
Watkins JS and Kovach RL (1972) Apollo 14 active seismic experiment. *Science* 175: 1244–1245.
Watters T and Johnson CL (2010) *Lunar tectonics*. *Planetary Tectonics*. Cambridge: Cambridge University Press.
Weaver PF, Yuen PC, Prolss GW, and Furumoto AS (1970) Acoustic coupling in the ionosphere from seismic waves of the earthquake at Kurile Islands on August 11, 1969. *Nature* 226: 1239–1241.
Webb SC (2007) The Earth's "hum" is driven by ocean waves over the continental shelves. *Nature* 445: 754–756.
Weber J (1960) Detection and generation of gravitational waves. *Physical Review* 117(1): 306–313.
Weber J (1971) Gravitational radiation experiments. In: Kuper CG and Peres A (eds.) *Relativity and Gravitation*. Proceedings of an International Seminar on Relativity and Gravitation, p. 309. New York: Gordon and Breach.
Weber RC, Bills BG, and Johnson CL (2009) Constraints on deep moonquake focal mechanisms through analyses of tidal stress. *Journal of Geophysical Research* 114: E05001. http://dx.doi.org/10.1029/2008JE003286.
Weber RC, Lin PY, Garnero EJ, Williams Q, and Lognonné P (2011) Seismic detection of the lunar core. *Science* 331: 309–312.
Widmer R and Zürn W (1992) Bichromatic excitation of long-period Rayleigh and air waves by the Mount Pinatubo and El Chichon volcanic eruptions. *Geophysical Research Letters* 19: 765–768.
Wieczorek MA, Bradley M, Jolliff L, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364.
Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339: 671–675. http://dx.doi.org/10.1126/science.1231530.
Williams JG, Boggs DH, Yoder CF, Ratcliff JT, Todd J, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
Withers MM, Aster RC, Young CJ, and Chael EP (1996) High-frequency analysis of seismic background noise as a function of wind speed and shallow depth. *Bulletin of the Seismological Society of America* 86(5): 1507–1515.
Wolf AA, Freeman A, Kedar S, Webb FH, and Matousek S (2012) Impactor science at Mars. In: *Concepts and Approaches for Mars Exploration*, Houston, TX, 12–14 June 2012. LPI, 1679, id. 4110.
Woodhouse JH and Dziewonski AM (1984) Mapping of the upper-mantle: Three dimensional modeling of Earth structure by inversion of seismic waveforms. *Journal of Geophysical Research* 89: 5953–5986.
Yamada R, Garcia RF, Lognonné P, Le Feuvre M, Calvet M, and Gagnepain-Beyneix J (2011) Optimisation of seismic network design: Application to a geophysical international lunar network. *Planetary and Space Science* 59: 343–354.
Yamada R, Garcia RF, Lognonné P, et al. (2013) On the possibility of lunar core phase detection using new seismometers for soft-landers in future lunar missions. *Planetary and Space Science* 81: 18–31.
Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303.
Yuen PC, Weaver PF, Suzuki RK, and Furumoto AS (1969) Continuous traveling coupling between seismic waves and the ionosphere evident in May 1968 Japan earthquake data. *Journal of Geophysical Research* 74: 2256–2264.
Zhao D, Arai T, Liu L, and Ohtani E (2012) Seismic tomography and geochemical evidence for lunar mantle heterogeneity: Comparing with Earth. *Global and Planetary Change* 90: 29–36.
Zhao D, Lei J, and Liu L (2008) Seismic tomography of the Moon. *Chinese Science Bulletin* 53: 3897–3907.
Zharkov VN and Gudkova TV (1997) On the dissipative factor of Martian interiors. *Planetary and Space Science* 45: 401–407.

Zharkov VN and Gudkova TV (2000) Interior structure models, Fe/Si ratio and parameters of figure for Mars. *Physics of the Earth and Planetary Interiors* 117: 407–420.

Zharkov VN and Gudkova TV (2005) Construction of Martian interior model. *Solar System Research* 39: 343–373. http://dx.doi.org/10.1007/s11208-005-0049-7.

Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 220–227.

Zuber MT, Smith DE, Watkins MM, et al. (2013) Gravity field of the Moon from the Gravity Recovery and Interior Laboratory (GRAIL) mission. *Science* 339: 668–671. http://dx.doi.org/10.1126/science.1231507.

Zuber MT, Solomon SC, Phillips RC, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

Zürn W and Widmer R (1996) World wide observation of bichromatic long-period Rayleigh-waves excited during the June 15, 1991 eruption of Mt. Pinatubo. In: Newhall C and Punongbayan R (eds.) *Fire and Mud, Eruptions of Mount Pinatubo, Philippines*, pp. 615–624. Quezon City/Seattle, WA: Philippine Institute of Volcanology and Seismology/University of Washington Press.

10.04 Rotation of the Terrestrial Planets

T Van Hoolst, Royal Observatory of Belgium, Brussels, Belgium; KU Leuven, Leuven, Belgium

© 2015 Elsevier B.V. All rights reserved.

10.04.1	**Introduction**	121
10.04.2	**Theoretical Foundations**	123
10.04.2.1	Conservation of Angular Momentum	123
10.04.2.1.1	Precession	123
10.04.2.1.2	Nutation, polar motion, and length-of-day variations	124
10.04.2.1.3	Kinematic relation between polar motion and nutation	125
10.04.2.2	Liouville Equations	126
10.04.2.2.1	Excitation	126
10.04.2.2.2	Basic equations	126
10.04.2.2.3	Core–mantle coupling	127
10.04.2.2.4	Mass redistribution	128
10.04.2.2.5	Periodic rotation variations	128
10.04.2.3	Hamiltonian Equations and Long-Term Evolution	129
10.04.2.4	Dissipative Effects as Causes of Secular Rotation Variations	131
10.04.2.4.1	Gravitational tides	131
10.04.2.4.2	Thermal atmospheric tides	132
10.04.2.4.3	Core–mantle friction	132
10.04.2.4.4	Climate friction	133
10.04.3	**Mars**	133
10.04.3.1	Current Rotation State	133
10.04.3.2	Long-Term Spin Evolution	133
10.04.3.3	Nutation	134
10.04.3.3.1	Rigid nutations	134
10.04.3.3.2	Nonrigid nutations	135
10.04.3.4	Rotation Speed Variations	136
10.04.3.5	Polar Motion	138
10.04.3.5.1	Seasonal polar motion	138
10.04.3.5.2	Chandler wobble	139
10.04.4	**Venus**	139
10.04.4.1	Current Rotation State	139
10.04.4.2	Long-Term Spin Evolution	140
10.04.4.3	Wobble, Length-of-Day Variations, and Nutation	140
10.04.5	**Mercury**	141
10.04.5.1	Current Rotation State	141
10.04.5.2	Long-Term Spin Evolution	141
10.04.5.2.1	Spin–orbit resonance	141
10.04.5.2.2	Precession about the Cassini state and nutation	143
10.04.5.3	Librations	144
10.04.5.3.1	Forced libration	144
10.04.5.3.2	Free libration	146
10.04.5.4	Wobble	146
10.04.6	**Summary**	147
Acknowledgments		147
References		147

10.04.1 Introduction

The rotation of the terrestrial planets of our solar system is remarkably diverse (see Table 1). The Earth and Mars spin much faster than their orbital motion, whereas the rotation periods of Venus and Mercury are commensurate with their orbital periods. The two innermost planets have very small obliquities (the angle between the orbital plane and the equator of the planet) as opposed to the large obliquities of about 25° for the Earth and Mars. Mercury, the Earth, and Mars have a prograde rotation, that is, their rotational motion is in the same sense as their orbital motion; Venus, however, has a retrograde rotation. Mercury is the only solar system planet in a spin–orbit resonance: its rotation period is two-thirds of its orbital period.

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00168-8

Table 1

Property	*Mercury*	*Venus*	*Earth*	*Mars*
M ($\times 10^{24}$ kg)	0.3302	4.8685	5.9736	0.6419
R (km)	2439	6052	6371	3390
a (Au)	0.38710	0.72333	1.00000	1.52366
e	0.20563	0.00677	0.01671	0.09341
ε (°)	2.04 ± 0.08 arcmin	177.4	23.44	25.19
P_{rot} (day)	58.646	243.020	0.997	1.026
P_{orb} (day)	87.969	224.695	365.256	686.980
C_{22}	$8.09 \pm 0.06 \times 10^{-6}$	$5.537 \pm 0.006 \times 10^{-7}$	1.574×10^{-3}	5.463×10^{-5}
J_2	$5.03 \pm 0.02 \times 10^{-5}$	$4.404 \pm 0.002 \times 10^{-6}$	1.083×10^{-3}	1.957×10^{-3}
$I/(MR^2)$	0.346 ± 0.014		0.3308	0.3645 ± 0.0005
$\dot{\psi}$ (arcsec year^{-1})	~2330*	~45*	50.288	7.594 ± 0.01

Mass M, mean radius R, semimajor axis a, eccentricity e, obliquity ε, siderial rotation period P_{rot}, orbital period P_{orb}, gravitational coefficients J_2 and C_{22}, normalized moment of inertia, and precession rate for the four terrestrial planets. Orbital elements are for J2000. See **Sections 10.04.3, 10.04.4**, and **10.04.5** for more information and references. Gravitational coefficients are from Smith et al. (2012) for Mercury, Konopliv et al. (1999) for Venus, and Konopliv et al. (2011) for Mars. A star indicates that the value has not been observed but has been calculated theoretically.

Planets accumulate rotational angular momentum from the relative motion of planetesimals at collision. The angular momentum acquired in an individual impact can be large and in any direction. However, there is a high degree of cancellation for the small planetesimals and the net rotation depends on a systematic bias in the distribution of the obtained angular momentum. For a realistic distribution of planetesimal eccentricities, planets accreting from a uniform surface density disk of small planetesimals rotate progradely although much slower than the Earth and Mars (Dones and Tremaine, 1993; Lissauer and Kary, 1991; Lissauer et al., 1997; Ohtsuki and Ida, 1998). A planet can acquire faster rotation in a nonuniform disk (see, e.g., Lissauer et al., 2000). On the other hand, during the late stages of terrestrial planet accumulation, one or a few large impacts can cause a planet to rotate very rapidly at nearly the breakup rate, so that the rotation of the planet is expected to be dominated by those random impacts. Large impacts can also determine the initial obliquity of a terrestrial planet. Due to the dominance of random phenomena and insufficient understanding of planetary accretion, definitive predictions regarding the accumulation of rotational angular momentum for each individual terrestrial planet cannot be made. Nevertheless, if giant impacts are solely responsible for the final spin state after planet formation, both the spin rates and obliquities should be isotropically distributed and prograde and retrograde spins are equally possible (Agnor et al., 1999; Chambers, 2001; Kokubo and Ida, 2007). Schlichting and Sari (2007) suggested that a final preference for prograde rotation results from the systematic prograde rotation of each protoplanet and from the accretion of remaining planetesimals in the post-giant impact phase.

In this chapter, we focus on the variations in the rotation of the terrestrial planets on short timescale and review how the rotation could have evolved from their primordial rotation state. We emphasize the broad geophysical and planetological significance of rotation studies and show that they provide a direct and unique tool to explore otherwise difficult to obtain global properties of the terrestrial planets. Rotation studies used to belong to the realm of celestial mechanics and the planets were usually considered to show a rigid rotational behavior. However, the internal structure and dynamics of a planet can have a large impact on its rotation. Also, the atmosphere, through exchanges of angular momentum with the solid planet, can cause the rotation of a planet to change. Here, we review rotation studies from a geophysical point of view with only limited information on the celestial mechanics. Tidal friction is a classical example of a geophysical effect on planetary rotation and the main cause of the secular braking of the rotation of the Earth. Dissipation of tidal energy in a planet delays the time of high tide, and the gravitational torque from the Sun on the tidal bulge – the tidal torque – changes the obliquity and tends to drive the planet to an equilibrium rotation rate, which is synchronous with its orbital motion for a circular orbit. Other dissipative effects, such as dissipative core–mantle interactions and thermal tides of the atmosphere, can also contribute substantially to the spin evolution of the planets. Also fluid layers, such as an atmosphere and a liquid core, can significantly change the rotation of the mantle and the crust on short timescales because they can rotate differentially with respect to the mantle and crust and interact with the solid layers. As an example, short periodic rotation variations as nutation and libration depend sensitively on the state of the core of the planet and can inform on core properties. The determination of the present core state by rotational means has profound implications for our understanding of the thermal evolution and the magnetic field history of the planet, as well as for the composition and thermal state at the time of formation.

Accurate present mean rotation states have been derived from Earth-based radar observations and space missions visiting the terrestrial planets. Periodic rotation variations are more difficult to measure, and only the largest components have been obtained with limited precision. More accurate data on all aspects of rotation variations and a full exploitation of the potential of rotation studies are foreseen with improved ground-based radar observations, upcoming orbiting spacecraft missions (such as the ESA BepiColombo mission to Mercury), and plans for radio science experiments with landers on the surface of terrestrial planets (such as with the NASA InSight mission to Mars).

The outline of this chapter is as follows. In **Section 10.04.2**, we introduce the basic concepts and equations, which are applied to Mars, Venus, and Mercury in **Sections 10.04.3,**

10.04.4, and 10.04.5, respectively. We start Section 10.04.2 with the equation of conservation of angular momentum and discuss properties of precession, nutation, polar motion, and rotation rate variations. We review in some detail the Liouville equation approach for the study of rotation variations and also briefly describe the Hamiltonian method for the case of rotation variations on long timescales. In the last part of Section 10.04.2, an overview is given of the secular effects of dissipative processes on the rotation of terrestrial planets. After an introduction on the current rotation state of Mars and its long-term variations, we present in Section 10.04.3 an overview of the nutation, length-of-day (LOD) variations, and polar motion of Mars in relation with its interior structure and atmosphere. The evolution of the rotation of Venus to its present state is reviewed in Section 10.04.4, together with its periodic rotation variations. In Section 10.04.5, we present an overview of the evolution and capture into resonance of Mercury and review its librations. In particular, we explain that accurate measurements of the librations and of the obliquity of Mercury put strong constraints on its interior structure.

10.04.2 Theoretical Foundations

10.04.2.1 Conservation of Angular Momentum

10.04.2.1.1 Precession

Rotation variations of the terrestrial planets are usually studied in a Newtonian approach. General relativistic effects are small because the squared velocities of the planetary motions are much smaller than the squared speed of light. Moreover, the ratio of the Schwarzschild radius to the radius of the planets is extremely small (for the Earth, about 1.4×10^{-9}). For the rotation variations of the Earth, the general relativistic effects have been shown to be negligibly small, except for an additional contribution to precession (e.g., Bizouard, 1995). This so-called geodetic precession, though larger than the observational accuracy, can nevertheless not be separated from Newtonian precession because of uncertainties in the flattening of the Earth. Since the general relativistic effects on rotation variations are expected to be equally small for the other terrestrial planets and the observational accuracies are a few orders of magnitude larger than for the Earth, we neglect the general relativistic effects here. In this approximation, we describe the rotation variations with respect to 'inertial reference frames' with origin in the mass center of the planets.

The fundamental equation governing the rotation of a terrestrial planet is the conservation equation for angular momentum valid in an inertial reference frame:

$$\frac{\mathrm{d}\vec{H}}{\mathrm{d}t}=\vec{L} \qquad [1]$$

Here, t is time, $\vec{H}$ is the spin angular momentum of the planet, and $\vec{L}$ is the external torque acting on it. If no external torque is applied on the planet, eqn [1] shows that the angular momentum will not change in space. However, other solar system bodies exert a gravitational torque on the nonspherically symmetrical terrestrial planets. This torque can be derived from the gravitational potential $V\left(\vec{r}\right)$ of the planet at the position $\vec{r}\,(r,\theta,\lambda)$ of the mass center of a perturbing solar system body, which can generally be expressed as

$$V\left(\vec{r}\right)=-\frac{GM}{r}\sum_{l}\sum_{m=0}^{l}\left(\frac{R_\mathrm{e}}{r}\right)^{l}[C_{lm}\cos(m\lambda) + S_{lm}\sin(m\lambda)]P_{lm}(\cos\theta) \qquad [2]$$

where M is the mass of the planet, R_e the mean equatorial radius of planet, θ the colatitude (90° declination), and λ the longitude (hour angle) of the perturbing body. The origin of the inertial reference frame is chosen in the mass center of the planet. The universal gravitational constant is denoted by G, and l and m are the degree and the azimuthal order of the associated Legendre function P_{lm}. Degree-one terms are zero because the origin of the coordinate system is chosen to be at the center of mass of the planet (Heiskanen and Moritz, 1967). The torque $\vec{L}$ exerted by a solar system body with mass M^* at position $\vec{r}$ on the planet can then be expressed as

$$\vec{L}=M^*\,\vec{r}\times\vec{\nabla}\,V \qquad [3]$$

As the magnitude of the terms in eqn [1] decreases rapidly with increasing degree l for bodies far from the planet ($r\gg Re$), only the second-degree terms are included, even in the most advanced and accurate rotation studies, except for the Earth–Moon system. In this approximation, the planet is thus considered to be a triaxial ellipsoid and the torques are proportional to the mass of the perturbing body and inversely proportional to r^3, where r is distance between the centers of mass of the two bodies. The torque is then often approximately written as

$$\vec{L}=3\frac{GMM^*}{r^3}(\hat{r}\times\mathbf{I}\cdot\hat{r}) \qquad [4]$$

(e.g., Bills, 2005a). Because the torque rapidly decreases with increasing distance from the planet, only the torques of the massive Sun and the natural satellites orbiting the planet must be considered.

When the planet is not in a spin–orbit resonance, the external solar torque averaged over the orbital period has only a component in the orbital plane perpendicular to the rotation vector (e.g., Chapter 3.10; Stacey, 1992). Therefore, both the norm of the angular momentum and its component in the direction perpendicular to the orbital plane are constant in time. As for a spinning top, the constant (time-independent) part of the gravitational torque then causes the angular momentum axis to precess about the normal to the orbit of the planet at constant obliquity. For a planet that is not in a spin–orbit resonance, the precession rate can be approximated by

$$\dot{\psi}=\alpha_\mathrm{P}\cos\varepsilon=-\frac{3}{2}\frac{n^2}{\Omega(1-e^2)^{3/2}}J_2\frac{MR^2}{C}\cos\varepsilon \qquad [5]$$

(see, e.g., Smart, 1953), where α_P is the precession constant, Ω the mean rotation rate of the planet, M the total mass, R the mean radius, n the mean (orbital) motion, e the eccentricity, and $J_2\equiv-C_{20}=(C-\overline{A})/MR^2$ the degree-two zonal gravity coefficient. The principal moments of inertia of the planet are denoted by $A<B<C$, and $\overline{A}=(A+B)/2$ is the mean equatorial moment of inertia. Obliquity ε is the angle between the equator and the orbital plane, and longitude ψ is an orientation angle of the angular momentum axis measured from the center of the circular path of precession (we assume here that the orbit

is fixed). A dotted quantity, such as $\dot{\psi}$, indicates the time derivative of that quantity. Interesting to note is that the precession rate (eqn [5]) only depends on the degree-two, order-zero coefficient of the gravitational potential of the planet.

For rapidly rotating planets (such as the Earth and Mars), the polar flattening is essentially due to rotation and proportional to Ω^2 (e.g., Jeffreys, 1952), and eqn [5] shows that the precession rate is proportional to the rotation rate. For planets rotating so slowly that the rotational flattening is only a minor contribution to their flattening (as is the case for Mercury and Venus), the torque is essentially independent of the rotation rate, while the spin angular momentum of the planet is linearly proportional to it. Therefore, the torque causes relatively larger changes in the angular momentum when the rotation rate decreases, such as for a spinning top, and the precession rate is inversely proportional to the rotation rate for slowly rotating planets.

The principal geophysical interest of precession is that it allows determining the polar principal moment of inertia C of the planet. Measurements of the gravitational field of the planet yield very accurate values for differences between principal moments of inertia, such as J_2, but are not sufficient to determine the individual principal moments of inertia (see Chapter 10.05). The mean moment of inertia of a planet gives insight into the mass concentration toward the center and is a major constraint for models of the interior structure of the terrestrial planets (Chapter 10.02).

10.04.2.1.2 Nutation, polar motion, and length-of-day variations

Because terrestrial planets move on elliptical orbits and have a nonzero obliquity and a nonspherically symmetrical form, the external gravitational torque from the Sun on them is time-dependent. Nutation of the angular momentum axis of a planet is the resulting small, time-dependent motion of that axis in space and can be expressed as a sum of periodic terms, whose main periods are subharmonics of the orbital period of the planet (for the Earth, see Chapter 3.10). Other planets and satellites can also exert a torque on the planet and cause additional nutational motion, but these nutations are generally negligible with respect to those related to the solar torque, except for the Earth's nutations due to the Moon.

Equation [1] shows that the change in angular momentum is entirely due to the external torque and further does not depend on the interior of the planet. This suggests that nutation observations would have no geophysical interest, other than providing a value for the relative moment of inertia difference (J_2/C), in addition to but less precise than precession (see eqn [5] and also eqns [65] and [66]). However, observations do not measure orientation changes of the angular momentum axis but of the figure axis or the rotation axis of the mantle. In studies of periodic rotation variations, three different polar axes are distinguished: the rotation axis, the figure axis, and the angular momentum axis. The figure axis is defined as the axis of greatest moment of inertia, and the rotation axis is the axis of rotation of the solid outer part of the planet, that is, the mantle and crust, if the (outer) core is liquid. In general, the rotation and figure axes do not coincide with the angular momentum axis, and their orientations do depend on the interior of the planet. We will see that the relation between both axes and the angular momentum axis depends in a complex way on the interior and that nutation observations give direct access to the properties of the interior of a terrestrial planet.

To analyze the nutation of figure and rotation axes, it is more convenient to rewrite eqn [1] in a reference frame attached to the rotating planet. Let $\vec{e}_1, \vec{e}_2, \vec{e}_3$ represent three orthonormal vectors along corotating, body-fixed axes, rotating with instantaneous angular velocity $\vec{\omega}$ with respect to an inertial reference frame. We will choose the rotation axis to be close to the third axis. For both the inertial and the corotating reference frames, the center of mass of the planet is chosen as the origin O. For any vector $\vec{V}$, the time derivative with respect to the corotating reference frame is related to the time derivative with respect to the inertial reference frame as

$$\left.\frac{\mathrm{d}\vec{V}}{\mathrm{d}t}\right|\text{inertial} = \left.\frac{\mathrm{d}\vec{V}}{\mathrm{d}t}\right|\text{co}-\text{rotating} + \omega\times\vec{V} \qquad [6]$$

(e.g., Goldstein, 1950), and conservation eqn [1] can be expressed as

$$\frac{\mathrm{d}\vec{H}}{\mathrm{d}t} + \vec{\omega}\times\vec{H} = \vec{L} \qquad [7]$$

where the time derivative is with respect to the body-fixed reference frame.

The angular momentum vector can be written as an integral over the total volume V of the planet as

$$\vec{H} = \int_V \vec{r}\times\vec{v}\,\rho\,\mathrm{d}V \qquad [8]$$

where $\vec{r}$ is the coordinate vector, $\vec{v}$ the velocity, and ρ the density of the mass elements. For a rigid planet, we have

$$\vec{H} = \int_V \vec{r}\times\left(\vec{\omega}\times\vec{r}\right)\rho\,\mathrm{d}V \qquad [9]$$

which can, by introducing the inertia tensor $\mathbf{I}$ with components

$$I_{ij} = \int_V \left[\left(x_1^2+x_2^2+x_3^2\right)\delta_{ij} - x_i x_j\right]\rho\,\mathrm{d}V \qquad [10]$$

be written as

$$\vec{H} = \mathbf{I}\cdot\vec{\omega}\,. \qquad [11]$$

Here, x_i are the rectangular Cartesian coordinates of the mass elements. By a suitable choice of coordinate axes, the inertia tensor can be brought to diagonal form and the angular momentum components reduce to

$$\left.\begin{aligned} H_1 &= A\omega_1,\\ H_2 &= B\omega_2,\\ H_3 &= C\omega_3. \end{aligned}\right\} \qquad [12]$$

For principal axis rotation, it follows that the figure axis and the rotation axis coincide with the angular momentum axis for a rigid planet.

By substituting eqn [12] into eqn [7], we obtain Euler equations for a rigid planet:

$$A\frac{\mathrm{d}\omega_1}{\mathrm{d}t} + (C-B)\omega_2\omega_3 = L_1 \qquad [13]$$

$$B\frac{d\omega_2}{dt}+(A-C)\omega_1\omega_3=L_2 \quad [14]$$

$$C\frac{d\omega_3}{dt}+(B-A)\omega_1\omega_2=L_3 \quad [15]$$

For rapidly spinning planets like the Earth and Mars, deviations from spherical symmetry are dominated by the flattening induced by rotation, which is symmetrical about the mean rotation axis x_3, so that the approximation $A=B$ can be used. For the Earth, for example, the relative difference between A and B is more than 100 times smaller than the relative difference between A and C. In the absence of external torques $\vec{L}$, the solution of Euler equations can then be written as

$$\left.\begin{aligned}\omega_1&=a\cos(\sigma_E t+f),\\ \omega_2&=a\sin(\sigma_E t+f),\\ \omega_3&=\Omega,\end{aligned}\right\} \quad [16]$$

where a is the amplitude and f the phase of the free polar motion. The motion is said to be free since both the amplitude and the phase depend only on the initial conditions if dissipation is neglected. Given an initial deviation of the rotation axis from the figure axis, the rotation axis rotates, or wobbles, about the figure axis with the Euler frequency:

$$\sigma_E=\frac{C-A}{A}\Omega. \quad [17]$$

This Euler wobble is a manifestation of the conservation of angular momentum when the figure axis and the rotation axis do not coincide. Solution (Borderies, 1980) is an example of wobble or polar motion, which in general describes the orientation variations of the rotation axis with respect to a body-fixed polar axis (for more precise definitions of polar motion and wobble, see **Chapters 3.10** and **3.09**). Polar motion can also be forced by motion in fluid planetary layers, such as an atmosphere and oceans, by redistributing mass and carrying additional angular moment (see **Section 10.04.2**). These motions can also cause periodic variations in the ω_3-component of the rota vector or changes in the rotation speed. As these variations imply changes in the length of the day, they are, in the geophysical literature, mostly referred to as length-of-day variations (ΔLOD).

10.04.2.1.3 Kinematic relation between polar motion and nutation

To describe rotation variations in space, the body-fixed reference frame must be related to an inertial reference frame, which we here take to have coordinates x_1', x_2', x_3' and an origin at the mass center of the planet. The transformation is usually described in terms of three rotations over three Euler angles (see **Figure 1**). We introduce the Euler angle ψ from the x_1'-axis to the intersecting line (ascending node line) between the (x_1', x_2')-plane of the inertial reference frame and the equator of the body-fixed reference frame, the Euler angle ε from the inertial x_3'-axis to the normal to the planet's equator (x_3-axis), and the Euler angle ϕ from the node line to the x_1-axis. If the orbital plane is assumed to be fixed and to coincide with the (x_1', x_2')-plane, ψ is the precession angle, ε the obliquity, and ϕ the rotation angle. The time derivatives of the Euler angles represent combinations of rotations of the three body-fixed axes with respect to the inertial frame. The relation between

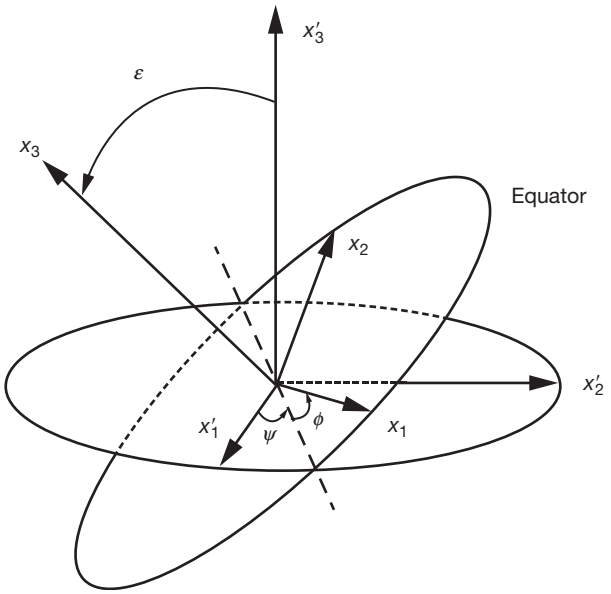


Figure 1 Euler angles.

the components (in the body frame) of the rotation vector and the time derivatives (in inertial space) of the Euler angles is given by Euler kinematic equations (e.g., Goldstein, 1950; Moritz and Mueller, 1987):

$$\left.\begin{aligned}\omega_1&=-\dot{\varepsilon}\cos\phi-\dot{\psi}\sin\varepsilon\sin\phi,\\ \omega_2&=\dot{\varepsilon}\sin\phi-\dot{\psi}\sin\varepsilon\cos\phi,\\ \omega_3&=\dot{\phi}+\dot{\psi}\cos\varepsilon.\end{aligned}\right\} \quad [18]$$

Because the figure axis is defined in terms of the planetary mass elements that rotate about the rotation axis, a variation in the rotation axis with respect to an inertial reference frame also leads to a variation in the figure axis with respect to the inertial reference frame. Moreover, these variations are not the same, and therefore, any periodic variation in the orientation of the rotation or figure axis in space (nutation) is accompanied by a periodic variation of the rotation axis with respect to the figure axis (polar motion or wobble), and vice versa. Consider, for example, the Euler wobble (eqn [16]). During the wobble, both the figure and the rotation axis have to change their orientation in inertial space; otherwise, the planet's spin angular momentum (which is defined in terms of the inertia and rotation of the planet; see eqn [11]) cannot be conserved. Mathematically, from the first two Euler kinematic equations (Bouquillon and Souchay, 1999), we have

$$\dot{\varepsilon}+i\sin\varepsilon\dot{\psi}=-\tilde{\omega}e^{i\phi} \quad [19]$$

where $\tilde{\omega}=\omega_1+\omega_2$ is the complex equatorial component of the rotation. If the x_3'-axis is chosen normal to the ecliptic at the time considered (the ecliptic of date) of the planet and the x_1'-axis points to the ascending node, $\dot{\varepsilon}$ and $\dot{\psi}$ represent changes in obliquity and in longitude (nutation). Equation [19] relates wobble described by the angle $\tilde{\omega}/\Omega$ to the time derivative of the nutation of the figure axis given by $\dot{\varepsilon}+i\sin\varepsilon\dot{\psi}$.. When the time dependence of ϕ is approximated by Ωt, periodic changes in the rotation vector with any forcing frequency σ in the corotating reference frame (wobble) are seen

to correspond to changes in longitude and obliquity (nutation) with frequency $\sigma' = \sigma + \Omega$ in the inertial reference frame.

By integrating eqn [19] with respect to time, it also follows that the amplitude of nutation of the figure axis differs a factor Ω/σ' from the amplitude of the wobble $(\tilde{\omega}/\Omega)$ of the rotation axis about the planet's figure axis. The motion of the rotation axis in space is the sum of the motion of the rotation axis about the figure axis (in Euler angles, $\delta_\varepsilon + i\sin\varepsilon\delta\psi = -i\tilde{\omega}\exp(i\Omega t)/\Omega$; see, e.g., Moritz and Mueller, 1987) and the motion of the figure axis in space. It then follows that the ratio of the amplitude of the wobble and the amplitude of nutation of the rotation axis is equal to σ'/σ. For the rapidly rotating planets, such as the Earth and Mars, the rotation period is much shorter than the orbital period. The main periods of nutation are related to the orbital motion and are therefore long compared to the rotation period $(\sigma'/\Omega \ll 1)$. Wobble periods are related to the seasons (and thus also to the orbital motion) and are therefore also long $(\sigma/\Omega \ll 1)$. The main nutations are therefore associated with very small wobble of diurnal period in the corotating reference frame, and large polar motion is associated with very small nutation of the rotation axis that is of diurnal period in space. When observations have limited time resolution, averaged orientations can only be obtained, and the small diurnal contributions to polar motion and nutation are neglected. For the Earth, the convention is to introduce an intermediate polar axis that only has low-frequency nutational motion in space and to define polar motion as the motion of that axis in the terrestrial reference frame outside the retrograde diurnal frequency band (e.g., Chapter 3.10). Each phenomenon can then be studied separately. Polar motion has been described before as the motion of the rotation axis about the figure axis, but actually, we see that the rotation axis remains nearly fixed with respect to inertial space. It is the figure axis, or the planet itself, that wobbles about the rotation axis.

10.04.2.2 Liouville Equations

10.04.2.2.1 Excitation

Both external and internal processes can excite periodic rotation variations of a terrestrial planet. We have sufficient knowledge about the external gravitational causes, but other processes are more difficult to model. For example, the seasonal global CO_2 exchange between the Martian atmosphere and the polar caps causes variations in the length of day of Mars, but is not precisely known. By studying rotation variations of a terrestrial planet, more can thus be learned about the excitation processes. Moreover, as the rotational response depends on the planet's structure and composition, also insight into the planetary interior can be obtained. This is particularly so for the rotation variations due to well-known external gravitational causes, such as for the nutations of Mars and the libration of Mercury.

10.04.2.2.2 Basic equations

Basic to most geophysical studies of the rotation of the Earth are the Liouville equations, which describe the change in angular momentum $\vec{H}$ of the deformable planet due to an applied torque (see, e.g., Chapter 3.10). These equations are also an ideal tool for investigating the rotation of the other terrestrial planets. We use a corotating reference frame as before. As a terrestrial planet is not rigid, we consider the three coordinate axes to be along the mean axes of principal moments of inertia of the mantle and to rotate with instantaneous angular velocity $\vec{\omega}$ with respect to an inertial reference frame. In equilibrium, the planet is assumed to rotate uniformly with angular velocity $\vec{\omega} = \Omega\vec{e}_3$.. Due to internal and external excitations, the rotation velocity of the mantle

$$\vec{\omega} = \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + , m_3 \end{pmatrix} \qquad [20]$$

deviates from the equilibrium rotation velocity aligned with the polar axis. The small deviations m_1 and m_2 describe wobble, and m_3 gives LOD variations.

In the reference frame rotating with the instantaneous rotation vector of the mantle, the change in angular momentum of the planet is governed by eqn [7]. The angular momentum (eqn [8]) can be expressed as

$$\vec{H} = \mathbf{I}\cdot\vec{\omega} + \vec{h} \qquad [21]$$

where

$$\mathbf{I} = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} + (c_{ij}). \qquad [22]$$

Mass redistribution and deformation of the planet cause small increments in inertia c_{ij}, and the small relative angular momentum $\vec{h}$ is due to relative motion with respect to the corotating frame in the fluid parts of the planet (liquid core, oceans, and atmosphere). We assume here that there is no rotation in the deformation field of the mantle $\left(\vec{h}^{\text{mantle}} = 0\right)$ with respect to the reference frame of the planet. The axes are then called mantle Tisserand axes (Moritz and Mueller, 1987; Munk and MacDonald, 1960). Equation [21] shows that part of the angular momentum is absorbed in relative motion, such as relative motion of a liquid core, and in perturbations of the moment of inertia. Rotation variations for a deformable planet are therefore different from those of a rigid planet. Deformations tend to decrease the rotation variations. A rotationally decoupled liquid core, on the other hand, may lead to larger mantle rotation variations because the mantle has a smaller moment of inertia than the total planet. By using eqns [20]–[22], the three components of the Liouville equations for the whole planet can be expressed as

$$A\Omega\dot{m}_1 + \Omega^2(C-B)m_2 = L_1 - \Omega\dot{c}_{13} + \Omega^2 c_{23} - \dot{h}_1 + \Omega h_2 \qquad [23]$$

$$B\Omega\dot{m}_2 + \Omega^2(C-A)m_1 = L_2 - \Omega\dot{c}_{23} + \Omega^2 c_{13} - \dot{h}_2 + \Omega h_1 \qquad [24]$$

$$C\Omega\dot{m}_3 = L_3 - \Omega\dot{c}_{33} - \dot{h}_3 \qquad [25]$$

These equations are linearized in m_i and c_{ij}, as these quantities are very small compared to their leading order terms in the equations. A particular consequence of this approximation is that there is no term in B – A in eqn [25], because that term is of second order in the rotation variations (e.g., Hinderer et al., 1982). Furthermore, the first two equations decouple from the third equation. Therefore, wobble and LOD variations are decoupled and can be calculated separately. The symmetry between eqn [25] on the one hand and eqns [23] and [24]

on the other hand is broken because the planet's mean rotation introduces a preferred direction, which we have chosen to be along the x_3-axis.

The right-hand sides of eqns [23]–[25] describe the excitations: external torques (Agnor et al., 1999), mass redistribution (Aharonson et al., 2004), and relative motion (Anderson, 1989). Of these, the external torque can most precisely be determined. The mass redistribution and relative motion on the other hand do depend not only on the excitation process but also on the rotation variations, as we will see later. The right-hand sides of eqns [23]–[25] divided by $\Omega\ (C-A)$ are called the excitation functions (Munk and MacDonald, 1960).

For rapidly rotating planets, the equatorial flattening (due to nonhydrostatic effects) is much smaller than the polar flattening (mainly due to the fast rotation), so the approximation $A=B$ can be used in the Liouville equations. By introducing the complex notation

$$\begin{aligned}\tilde{m}&=m_1+\mathrm{i}m_2\\ \tilde{c}&=c_{13}+\mathrm{i}c_{23}\\ \tilde{h}&=h_1+\mathrm{i}h_2\end{aligned} \qquad [26]$$

wobble can be described by the complex sum of the first two components of the Liouville eqns [23] and [24]:

$$A\Omega\dot{\tilde{m}}-\mathrm{i}(C-A)\Omega^2\tilde{m}=-\Omega\dot{\tilde{c}}-\mathrm{i}\Omega^2\tilde{c}+\dot{\tilde{h}}-\mathrm{i}\Omega\tilde{h} \qquad [27]$$

If the planet's core is liquid, Hough (1895) (see also Poincaré, 1910) showed with mathematical rigor that a change in the mantle rotation leads to a change in the core flow that can be expressed as a sum of two terms: a rigid body rotation of the core, which is of the same order of magnitude as the mantle rotation change, and an irrotational core flow, which is one order in the core flattening smaller. This result was derived for an incompressible, inviscid, and homogeneous core, but further studies showed it to be a valid approximation when these conditions are relaxed (see, e.g., Busse, 1968; Roberts and Stewartson, 1965; Stewartson and Roberts, 1963 for viscosity, and Smith, 1977 for a numerical investigation of the problem).

Three additional quantities m_i^{f}, describing the relative rotation of the fluid core with respect to the mantle, have to be determined when the planet's core is liquid. We define these by

$$\vec{\omega}^{\mathrm{f}}=\Omega\begin{pmatrix}m_1^{\mathrm{f}}\\ m_2^{\mathrm{f}}\\ m_3^{\mathrm{f}}\end{pmatrix}. \qquad [28]$$

The total rotation vector of the core is then $\vec{\omega}+\vec{\omega}^{\mathrm{f}}$. It is assumed here that the core corotates with the mantle in the equilibrium state. The differential rotations of the core are determined by three additional Liouville equations, which express the rate of change of angular momentum of the core as a result of the torque $\vec{L}$ exerted on the core in the same mantle corotating reference frame as before:

$$\frac{\mathrm{d}\vec{H}^{\mathrm{f}}}{\mathrm{d}t}+\vec{\omega}\times\vec{H}^{\mathrm{f}}=\vec{L}^{\mathrm{f}} \qquad [29]$$

where

$$\vec{H}^{\mathrm{f}}=\mathbf{I}^{\mathrm{f}}\cdot\vec{\omega}+\vec{h}^{\mathrm{f}} \qquad [30]$$

$$\begin{aligned}&=\Omega\left[\begin{pmatrix}A^{\mathrm{f}}&0&0\\0&B^{\mathrm{f}}&0\\0&0&C^{\mathrm{f}}\end{pmatrix}+\begin{pmatrix}c_{ij}^{\mathrm{f}}\end{pmatrix}\right]\\ &\times\begin{pmatrix}m_1\\ m_2\\ 1+m_3\end{pmatrix}+\begin{pmatrix}A^{\mathrm{f}}m_1^{\mathrm{f}}\\ B^{\mathrm{f}}m_2^{\mathrm{f}}\\ C^{\mathrm{f}}m_3^{\mathrm{f}}\end{pmatrix},\end{aligned} \qquad [31]$$

and $A^{\mathrm{f}}<B^{\mathrm{f}}<C^{\mathrm{f}}$ are the three principal moments of inertia of the fluid core and c_{ij}^{f} the small increments in the inertia of the core. The three components of the Liouville equations for the liquid core can then be written as

$$\begin{aligned}&A^{\mathrm{f}}\Omega\left(\dot{m}_1+\dot{m}_1^{\mathrm{f}}\right)-B^{\mathrm{f}}\Omega^2m_2^{\mathrm{f}}+\Omega^2\left(C^{\mathrm{f}}-B^{\mathrm{f}}\right)m_2+\Omega\dot{c}_{13}^{\mathrm{f}}-\Omega^2c_{23}^{\mathrm{f}}\\ &=L_1^{\mathrm{f}}\end{aligned} \qquad [32]$$

$$\begin{aligned}&B^{\mathrm{f}}\Omega\left(\dot{m}_2+\dot{m}_2^{\mathrm{f}}\right)-A^{\mathrm{f}}\Omega^2m_1^{\mathrm{f}}+\Omega^2\left(C^{\mathrm{f}}-A^{\mathrm{f}}\right)m_1+\Omega\dot{c}_{23}^{\mathrm{f}}-\Omega^2c_{13}^{\mathrm{f}}\\ &=L_2^{\mathrm{f}},\end{aligned} \qquad [33]$$

$$C^{\mathrm{f}}\Omega\left(\dot{m}_3+\dot{m}_3^{\mathrm{f}}\right)+\dot{c}_{33}^{\mathrm{f}}=L_3^{\mathrm{f}}, \qquad [34]$$

where only terms linear in m_i, m_i^{f}, c_{ij}, and c_{ij}^{f} were retained. As for the Liouville equations for the whole planet, the rotation rate variations are decoupled from the orientation variations, and there is no term in $B^{\mathrm{f}}-A^{\mathrm{f}}$ in eqn [34]. To obtain the solutions for wobble (m_x and m_y) and LOD variations (m_z) of a planet with a liquid core, the deformation-induced incremental inertia terms ($c_{13}, c_{23}, c_{13}^{\mathrm{f}}, c_{23}^{\mathrm{f}}, c_{33}, c_{33}^{\mathrm{f}}$), the relative angular momentum (h_1, h_2, h_3), and the torques ($L_1, L_2, L_3, L_1^{\mathrm{f}}$, and L_3^{f}) have to be determined from the excitation processes considered. For every frequency σ, the associated nutations of the rotation axis have an amplitude differing a factor σ/σ' as follows from integrating Euler's kinematic eqn [19] with respect to time.

10.04.2.2.3 Core–mantle coupling

The torque on the fluid core is the sum of the external gravitational torque on the permanent core figure due to the Sun and other solar system bodies and the torque due to the mantle. The latter torque includes viscous coupling between the mantle and the core, electromagnetic coupling, gravitational coupling, pressure coupling due to the ellipsoidal core form, and pressure coupling due to topography on the CMB (core–mantle boundary). By assuming the dissipative torques (viscous and electromagnetic) to be negligibly small for relatively short-period rotation variations (but see **Section 10.04.2.4** for their effect on the long-term rotation) and by neglecting CMB topography other than the ellipticity, the torque on the fluid core is often written as the sum of the external gravitational torque $\left(\vec{L}_{\mathrm{ext}}^{\mathrm{f}}\right)$ and the so-called inertial torque $\left(\vec{N}\right)$, which is the sum of the gravitational and pressure torque. By using the Navier–Stokes equation for the fluid core, the inertial torque can be calculated by assuming a Poincaré flow in the fluid core. Correct up to the first order in the flattenings and the rotation variations, we have

$$\begin{aligned}N_1&=-L_{\mathrm{ext},1}^{\mathrm{f}}+\Omega^2\left(C^{\mathrm{f}}-B^{\mathrm{f}}\right)\left(m_2+m_2^{\mathrm{f}}\right)-\Omega^2c_{23}^{\mathrm{f}}\\ N_2&=-L_{\mathrm{ext},2}^{\mathrm{f}}+\Omega^2\left(A^{\mathrm{f}}-C^{\mathrm{f}}\right)\left(m_1+m_2^{\mathrm{f}}\right)-\Omega^2c_{13}^{\mathrm{f}}\\ N_3&=-L_{\mathrm{ext},3}^{\mathrm{f}}\end{aligned} \qquad [35]$$

(e.g., Rochester, 1976; Sasao et al., 1977, 1980; and Hinderer et al., 1982; Van Hoolst and Dehant, 2002 for the triaxial case).

By substituting these expressions into Liouville eqns [32]–[34] for the core, we have

$$\frac{\mathrm{d}\vec{H}^{\mathrm{f}}}{\mathrm{d}t}-\vec{\omega}^{\mathrm{f}}\times\vec{H}^{\mathrm{f}}=0 \qquad [36]$$

(Mathews et al., 1991; Sasao et al., 1980). From the expression (Chapman and Lindzen, 1970) for the third component of the inertial torque, it follows that $L_3^{\mathrm{f}}=0$ and

$$C^{\mathrm{f}}\Omega\left(\dot{m}_3+\dot{m}_3^{\mathrm{f}}\right)+\dot{c}_{33}^{\mathrm{f}}\Omega=0. \qquad [37]$$

If deformation is neglected, there is no net change in the third component of the angular momentum of the core in the lowest-order approximation in the rotation variations and the flattenings: the external torque is balanced by the core–mantle coupling mechanisms for the third component. For an incompressible, inviscid, and homogeneous core, Hough (1895) already demonstrated that the additional core rotation with respect to the equilibrium rotation is purely equatorial, so that the rotation rate of the core remains constant. There is no complete torque balance for the equatorial components. Again, this symmetry breaking between the axial and the equatorial components is due to the mean rotation of the planet, which introduces a preferred direction in the problem.

10.04.2.2.4 Mass redistribution

We now discuss the problem of determining the incremental inertia terms. In particular, we consider the interesting case of excitation by an atmosphere as a typical example that includes excitation due to both relative motion and mass redistribution. By setting the external torque equal to zero and introducing the equatorial angular momentum function $\tilde{\chi}=\frac{\Omega\tilde{c}+\tilde{h}}{(C-A)\Omega}$ and the axial angular momentum function $\chi_3=\frac{\Omega c_{33}+h_3}{c\Omega}$, the Liouville equations [27] and [25] can be reexpressed as

$$\tilde{m}+\frac{\mathrm{i}}{\sigma E}\frac{\mathrm{d}\tilde{m}}{\mathrm{d}t}=\tilde{\chi}-\frac{\mathrm{i}}{\Omega}\frac{\mathrm{d}\tilde{\chi}}{\mathrm{d}t}, \qquad [38]$$

$$\dot{m}_3=-\dot{\chi}_3. \qquad [39]$$

The contribution of the atmosphere to the angular momentum functions can be separated in a term due to the wind ($\tilde{h}$ and h_3) and a term due to the atmospheric mass ($\tilde{c}$ and c_{33}). Explicit expressions for these 'wind' and 'matter' atmospheric angular momentum functions in terms of wind velocity and atmospheric pressure can be found in, for example, Barnes et al. (1983). For a rigid planet, only the atmospheric wind angular momentum functions, $\tilde{\chi}_{\mathrm{A}}^{\mathrm{W}}$ and $\chi_{3,\mathrm{A}}^{\mathrm{W}}$, and atmospheric matter angular momentum functions, $\tilde{\chi}_{\mathrm{A}}^{\mathrm{W}}$ and $\chi_{3,\mathrm{A}}^{\mathrm{M}}$, have to be considered in the Liouville eqns [38] and [39]. The superscript A denotes that only the direct contribution from the atmosphere (and seasonal polar caps) is considered, and superscripts M and W indicate matter and wind terms, respectively.

For an entirely solid but nonrigid planet, contribution to the angular momentum functions from the incremental inertia terms $\tilde{c}=c_{13}+\mathrm{i}c_{23}$ and c_{33} due to the planet's response has to be added. Two effects contribute to these inertia terms. First, when the rotation vector of the planet changes, the centrifugal potential changes and induces a slight mass redistribution of the planet. The associated changes in the inertia terms (indicated by superscript 'center') can for any planet be determined from McCullagh's theorem:

$$\tilde{c}^{\mathrm{center}}=k_2\frac{\Omega^2R^5}{3G}\tilde{m}, \qquad [40]$$

$$c_{33}^{\mathrm{center}}=k_2\frac{4R^5\Omega^2}{9G}m_3, \qquad [41]$$

where k_2 is the potential Love number of degree 2 of the planet (Moritz and Mueller, 1987; Munk and MacDonald, 1960). These contributions to the inertia terms are independent of the excitation source.

A second contribution to the incremental inertia terms $\tilde{c}$ and c_{33} is due to the atmosphere (and ice cap) loading. The loading induces a deformation and a corresponding change in the mass distribution of the planet. The associated change in the planetary gravitational potential at the planet's surface is most conveniently expressed as k_2' times the potential of the load, where k_2' is the degree-two load Love number (Munk and MacDonald, 1960). Since the incremental moments of inertia associated with the change in gravitational potential due to the atmospheric loading are linearly proportional to the change in the gravitational potential according to McCullagh's theorem, it follows that the contribution of the loading effect is $k_2'\tilde{\chi}_{\mathrm{A}}^{\mathrm{M}}$ to the equatorial angular momentum function and $k_2'\chi_{3,\mathrm{A}}^{\mathrm{M}}$ to the axial angular momentum function.

10.04.2.2.5 Periodic rotation variations

By taking the mass redistribution into account and solving eqns [38] and [39] at a given frequency σ, wobble of a deformable solid planet can be expressed as

$$\tilde{m}(\sigma)=-\frac{\alpha(\sigma+\Omega)}{(\sigma-\sigma c w)}\left[\left(1-k_2'\right)\tilde{\chi}_{\mathrm{A}}^{\mathrm{M}}(\sigma)+\tilde{\chi}_{\mathrm{A}}^{\mathrm{W}}(\sigma)\right], \qquad [42]$$

(e.g., Moritz and Mueller, 1987), and relative rotation rate variations are given by

$$m_3(\sigma)=-\left[\left(1+k_2'\right)\chi_{3,\mathrm{A}}^{\mathrm{M}}(\sigma)+\chi_{3,\mathrm{A}}^{\mathrm{W}}(\sigma)\right]\times\left(1+k_2\frac{4R^5\Omega^2}{9GC}\right)^{-1}. \qquad [43]$$

Here,

$$\alpha=\frac{C-\frac{A+B}{2}}{\frac{A+B}{2}} \qquad [44]$$

is the dynamic polar flattening. The only differences with respect to a solid planet are the term in the loading Love number k_2', and the change from the Euler frequency σ_{E} to the frequency σ_{CW} of the Chandler wobble (CW), which is the extension of the Euler wobble to a deformable planet (see also eqn [47] later). The difference between the CW frequency and the Euler frequency is due to the deformation of the planet as a result of the change in the centrifugal potential associated with the wobble.

For planets with a liquid core, the rotation rate variations differ only by a factor C/C^{m}, where C^{m} is the polar moment of inertia of the mantle, with respect to a solid planet:

$$m_3(\sigma)=-\frac{C}{C^{\mathrm{m}}}\left[\left(1+k_2'\right)\chi_{3,\mathrm{A}}^{\mathrm{M}}(\sigma)+\chi_{3,\mathrm{A}}^{\mathrm{W}}(\sigma)\right] \qquad [45]$$

when the small inertia terms due to the centrifugal force are neglected. These terms describe the additional variation in rotation rate due to the changes in the polar moment of inertia resulting from the rotation rate variations and are a second-order effect.

For polar motion of planets with a liquid core, the incremental inertia terms $\tilde{c}^{f} = c^{f}_{13} + \mathrm{i}c^{f}_{23}$ and c^{f}_{33} do have to be determined. The main reason is that for polar motion, changes in the inertia tensor are compared to differences in the moments of inertia (see eqns [23], [24], [32], and [33]), and for rotation rate variations, they are compared to the polar moment of inertia C itself (see eqns [25] and [34]), which is several orders of magnitude larger than the differences in principal moments of inertia for the terrestrial planets. The incremental core inertia terms can be expressed in terms of compliances or generalized Love numbers (**Chapter 3.10**; Dehant et al., 1993; Hinderer et al., 1982; Mathews et al., 1991; Sasao et al., 1977, 1980). By solving the Liouville equations for the entire planet and the liquid core, wobble can then be expressed as

$$\tilde{m}(\sigma) = -\frac{\alpha A(\sigma-\Omega)}{A^{m}(\sigma-\sigma cw)}\left(1+\frac{A^{f}}{A}\frac{\sigma'}{\sigma'-\sigma'_{FCN}}\right) \times\left[\left(1+k'_{2}\right)\tilde{\chi}^{M}_{A}(\sigma)\tilde{\chi}^{W}_{A}(\sigma)\right] \quad [46]$$

(Defraigne et al., 2000; Sasao et al., 1980), where A^{m} is the equatorial moment of inertia of the mantle with respect to the first body-fixed axis. Here, σ_{cw} and σ'_{FCN} are the frequencies of the CW and the free core nutation (FCN). These normal modes can be determined by solving the Liouville equations in the case of no internal or external excitation. The CW consists essentially of a rigid rotation of the mantle about the mantle rotation axis. The FCN is mainly a rigid rotation of the fluid core about the mean mantle rotation axis and only exists when the core is liquid. As they are defined in the absence of external excitation, the normal modes conserve the total spin angular momentum of the planet. Both the CW and the FCN have their origin in the flattening of the planet. The frequencies of the CW and the FCN can, correct up to the first order in the flattenings, be expressed as (**Chapter 3.10**; Sasao et al., 1980; Van Hoolst and Dehant, 2002; Yoder and Standish, 1997)

$$\sigma_{cw} = \Omega\frac{A}{A^{m}}\sqrt{\left(1-\frac{k_{2}}{k}\right)^{2}\alpha^{2}-\frac{\beta^{2}}{4}} \quad [47]$$

$$\sigma'_{FCN} = -\Omega(\alpha_{f}-\beta_{d})A/A_{m} \quad [48]$$

where $k = 3\,J_{2}/q_{0}$ is the secular Love number, α_{f} the dynamic polar flattening at the CMB, and β_{d} a compliance that describes the deformations due to the centrifugal effect of the differential wobble of the core. We also used the equatorial flattening

$$\beta = \frac{B-A}{A} \quad [49]$$

and the ratio of the surface centrifugal acceleration to the surface gravity

$$q_{0} = \frac{\Omega^{2}R^{3}}{GM}. \quad [50]$$

Note that the frequency of the FCN in eqn [48], as indicated by the prime, is expressed with respect to the inertial reference frame and that the CW frequency in eqn [47] is given in the body-fixed reference frame. As the flattenings are small for terrestrial planets, both these periods are long compared to the rotation period of the planet. Note, however, that the CW period is long in a corotating reference frame and that the FCN has a long period with respect to an inertial reference frame. Equations [47] and [48] show that deformation ($k_{2} > 0$ and $\beta_{d} > 0$) lengthens the CW and FCN periods and that a liquid core decreases the CW period.

Solutions [42] and [46] for wobble of a solid planet and a planet with a liquid core show resonances at the frequencies of the CW and the FCN. These normal modes of the planet are characterized by a predominant rotational motion and are sometimes called rotational normal modes. In general, the rotational response of a planet to an external forcing can be expressed as a linear combination of the planet's normal eigenfunctions, which form a complete set (Wahr, 1981a,b). Although the largest – and most familiar – class of normal modes is the set of seismic modes, the rotational normal modes are more important for the description and understanding of planetary rotation variations. The seismic modes have very small rotation in their displacement field and therefore cannot appreciably affect the rotation of a planet. They also have much smaller periods than the periods at which rotation variations are excited. For the Earth, the seismic mode with the longest period is the so-called football mode, with a period of 53.9 min (Dahlen and Tromp, 1998). Most other modes have periods of the order of minutes or tens of minutes. The three other terrestrial planets are smaller and less dense, and, being of similar structure, their seismic periods are smaller. As an example, for Mars, the football mode has a period of about 30 min (Gudkova and Zharkov, 2004; Van Hoolst et al., 2000a). As the rotational forcing frequencies are far from the frequencies of the seismic normal modes, there are no resonance effects with these modes.

Four rotational normal modes mainly related to orientation changes can be distinguished for terrestrial planets with at most three different internal layers that can show relative rotation (solid mantle + crust, liquid outer core, and solid inner core): the CW, the FCN, the inner core wobble (ICW), and the free inner core nutation (FICN). The ICW is analogous to the CW and describes the relative rotation of the figure axis of the inner core about the mantle rotation axis. The FICN represents the relative rotation of the inner core about the mantle rotation axis. The ICW has a long period in a corotating frame and the FICN a long period in an inertial frame. Extended Liouville theories for terrestrial planets with a solid elastic mantle, a liquid outer core, and an elastic inner core show that the solution for polar motion shows resonances at the eigenfrequencies of these four rotational normal modes (see, e.g., Dehant et al., 1993, 2005; Mathews et al., 1991, 2002). An important consequence of these resonances is that for frequencies close to the resonant frequencies, rotation variations can be amplified. The detection of the amplification yields information on the normal mode frequencies and thus on the internal properties on which these frequencies depend (see, e.g., eqns [47] and [48]).

10.04.2.3 Hamiltonian Equations and Long-Term Evolution

In astronomical studies of planetary rotation variations in which the planets are considered to be rigid (as opposed to

geophysical studies), Hamiltonian and Lagrangian approaches are customarily used (Kinoshita, 1977; Smart, 1953). These methods have also been extended to include effects of deformation and core–mantle coupling for the nutations of the Earth (see, e.g., Getino and Ferrándiz, 2000) but have not yet reached that level of precision for the short-term variations in the rotation of the other terrestrial planets. On the other hand, for studies of the long-term evolution of planetary spins, Hamiltonian and Lagrangian equations are often preferred to torque equations. Hence, we here briefly introduce the Hamiltonian approach to planetary rotation in the frame of long-term changes of rotation.

For the long-term rotational evolution, it is customary to neglect polar motion and nutation and to assume that the terrestrial planets rotate about their greatest principal inertia axis, corresponding to the lowest rotational energy, and that the angular momentum axis also coincides with that axis, as would be the case for a rigid planet with principal axis rotation. These assumptions are justified since polar motion and nutation are very small compared to the long-term changes (see **Sections 10.04.3, 10.04.4**, and **10.04.5**) and average to zero over a sufficiently long time. Moreover, large wobble at any moment in the history of a planet will be damped in a relatively short time compared to the age of the solar system (Peale, 1976a; Yoder and Ward, 1979). Nevertheless, the terrestrial planets most likely all have had periods of large polar wander when the solid surface of the planet reorients with respect to the rotation axis as a result of mass redistribution due to internal or external causes that drives the planet to a new state of minimum energy corresponding to principal axis rotation (Gold, 1955; Matsuyama et al., 2006; Willemann, 1984). For example, large polar wander of a few 10° is thought to be caused by the emplacement of Tharsis (Matsuyama and Manga, 2010; see also Zhong, 2009).

Basic to the Hamiltonian and Lagrangian methods is the development of expressions for the kinetic energy and potential energy of the rotating planet interacting with other solar system bodies. The rotational kinetic energy T of a planet supposed rigid is given by

$$T=\frac{1}{2}\left(A\omega_1^2+B\omega_2^2+C\omega_3^2\right) \qquad [51]$$

when planetary effects on the orbit are neglected (see, e.g., Goldstein, 1950). The components of the angular velocity $\vec{\omega}$ are defined with respect to the principal axes of inertia of the planet. The potential energy of an isolated rigid body does not change with time and can therefore be neglected. The only potential energy to be included then is that due to the gravitational interaction with the Sun, the other planets, and satellites. By retaining only the degree-two terms of eqn [2], the potential energy U for an interacting body with mass M^* can be expressed as

$$U=-\frac{GMM^*}{r}-\frac{GMM^*}{2r^3}(A+B+C-3\mathcal{I}), \qquad [52]$$

where $\mathcal{I}$ is the moment of the inertia of the planet with respect to the direction from the planet's mass center to the perturbing body.

In the Lagrangian equations,

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}_n}\right)-\frac{\partial\mathcal{L}}{\partial q_n}=0, \qquad [53]$$

where $\mathcal{L}=T-U$ is the Lagrangian and the Euler angles are taken as the generalized coordinates q_n. By using eqn [18], the kinetic energy can readily be expressed in terms of the time derivatives of the Euler angles. For Hamiltonian equations, the Euler angles and their adjoint moments could be used. However, these moments have no intuitive physical meaning, and often Andoyer variables are preferred (Andoyer, 1923; Kinoshita, 1977). For the case of principal axis rotation, the problem is reduced from six variables to four variables (Kinoshita, 1977). The two Andoyer variables to be used are (Agnor et al., 1999) the angular momentum amplitude $H=C\omega$ and (Aharonson et al., 2004) the angular momentum component normal to the orbital plane of date, $X=H\cos\varepsilon$, and the two conjugate Andoyer variables are (Agnor et al., 1999) the hour angle θ between the vernal equinox of date and a fixed direction in the equator (e.g., the axis of minimum moment of inertia) and (Aharonson et al., 2004) the precession ψ (in general, taking into account the motion of the orbit: see, e.g., Kinoshita, 1977). In addition to the hour angle, the Hamiltonian also depends on a second fast variable describing the revolution of the planet around the Sun. For studies of the long-term evolution of the planet, the Hamiltonian is averaged over these fast angles. For a rigid planet with principal axis rotation and without planetary perturbations and spin–orbit resonances, the resulting averaged Hamiltonian can be expressed as

$$\mathcal{H}_{\text{averaged}}=\frac{H^2}{2C}-\frac{\alpha_P}{2}\frac{X^2}{H} \qquad [54]$$

(see, e.g., Néron de Surgy and Laskar, 1997), where α_P is the precession constant (eqn [5]), and the Hamiltonian equations yield

$$\left.\begin{aligned}\frac{dH}{dt}&=0\\ \frac{dX}{dt}&=0\\ \frac{d\theta}{dt}&=\frac{H}{C}+\frac{a_P}{2}\frac{X^2}{H^2}\\ \frac{d\psi}{dt}&=\alpha_P\frac{X}{H}\end{aligned}\right\} \qquad [55]$$

Within the approximations made, it follows that the rotation speed of the planet, the obliquity, and the precession rate are constant in time.

In the case of spin–orbit resonances, that is, when the ratio of the rotation speed ω and the mean motion n approximates a rational number $p=\text{integer}/2$, combinations $\theta-pM$ of the two fast variables (M is the mean anomaly) appearing in the Hamiltonian cease to be fast variable and cannot be eliminated by the averaging procedure. The Andoyer variables H and X, or equivalently the rotation speed and the obliquity, then become time-dependent, with time derivatives being proportional to $\sin2(\theta-pM)$, and the spin–orbit resonances also contribute a term to the precession (see, e.g., Correia et al., 2003). For Mercury, the resulting changes in rotation speed (librations), obliquity variations, and precession rate will be studied in **Section 10.04.5**.

Additional effects can be taken into account by adding the corresponding energy to the Hamiltonian. Effects related to dissipative processes will be studied in **Section 10.04.2.4**. Here, we briefly consider the inclusion of planetary perturbations. Because the ecliptic of date, with respect to which the Andoyer variables are defined, changes its orientation due to planetary gravitational effects, additional terms have to be added to the Hamiltonian equations (eqn [55]) for X and ψ (e.g., Néron de Surgy and Laskar, 1997; Peale, 1973). The indirect effects of the other planets on the orbit are much more important than the direct planetary gravitational torque on the planet. As mentioned earlier, the direct torque is usually neglected, and so H remains constant. Planetary effects alone therefore do not affect the rotation rate of a planet on a long timescale, but can change the obliquity and precession. As will be discussed in **Section 10.04.5** on Mercury, planetary perturbations on the orbit can also periodically change the rotation rate on short timescales by indirectly modifying the solar torque.

If the spin axis of a terrestrial planet were fixed in inertial space, obliquity variations would be entirely due to the planetary-induced variations in the orientation of the orbital plane (called orbital precession as the normal to the orbital plane precesses about its mean orientation). However, the solar torque (and also the lunar torque for the Earth) on the flattened planet causes the planet to precess about the changing orbit normal. The resulting motion of the spin axis with respect to the orbit normal depends critically on the ratio of the precession period to the period of the orbital precession. For a very small ratio, the spin axis precession is about the slowly moving orbit normal and tracks that motion efficiently such that the obliquity is nearly constant. On the other hand, if the precession of the spin axis is much slower than the precession of the orbit, the spin axis will precess about the averaged orbit normal (see, e.g., Ward, 1992). If the ratio is about one (as is now the case for Mars), we have a resonant situation, and the obliquity variations can be large and chaotic (e.g., Touma and Wisdom, 1993; Ward, 1974). Laskar and Robutel (1993) have shown that the four terrestrial planets could have experienced large chaotic variations in obliquity during some period(s) of their history and that their obliquities cannot be considered as primordial.

10.04.2.4 Dissipative Effects as Causes of Secular Rotation Variations

10.04.2.4.1 Gravitational tides

Dissipation of tidal energy in the Earth is the principal cause of the secular increase in the length of day (Munk and MacDonald, 1960). Deviations from perfect elasticity of the planetary material delay the response of the planet to an applied tidal torque and cause the tidal bulge to be slightly misaligned with respect to the direction to the body raising the tides. The misalignment results in a small torque on the planet that alters the rotation of the planet. The theory of tides and the influence on the rotation and orbits of planets and satellites have been described in many books and review articles and will not be repeated here in detail (see, e.g., Darwin, 1908; Goldreich and Soter, 1966; Kaula, 1964; Lambeck, 1980; MacDonald, 1964; Melchior, 1983; Munk and MacDonald, 1960; Murray and Dermott, 1999). An important property is that the main part of the tidal potential is proportional to r^{-3} (see the degree-two part of eqn [2]), where r is the distance between the planet and the tide-raising body. Therefore, the tidal bulge is also proportional to r^{-3}. Since the gravitational torque on the planet is also proportional to r^{-3} (eqn [4]), the tidal torque is then proportional to r^{-6}. As for the Earth, the solar tidal torque slows down the rotation of a rapidly rotating terrestrial planet ($\omega > n$ or $\omega < -n$) since then the tidal bulge leads the planet–Sun direction when measured in the sense of rotation of the planet.

Because of the strong dependence on the distance to the Sun, tidal braking due to the Sun is much more efficient for Mercury and Venus than for the Earth and Mars and could be the cause of the slow rotation of Mercury and Venus compared to the fast rotation of the Earth and Mars. From eqn [1] and a tidal torque of magnitude

$$L_{\mathrm{T}} = \frac{3k_2 GM_{\odot}^2 R^5}{2a^6} \sin 2\delta \qquad [56]$$

for a circular orbit and zero obliquity (e.g., MacDonald, 1964; Murray and Dermott, 1999), the rate of decrease of rotation velocity can be approximated by

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -\frac{3k_2 GM_{\odot}^2 R^5}{2Ca^6 Q}, \qquad [57]$$

where $Q \approx 1/(2\delta)$ is the dissipation or quality factor of the planet and δ the geometric lag angle between the tidal bulge and the line planet–Sun (the factor 2 stems from the fact that the main tides are semidiurnal with frequency 2ω for rapid rotators). The despin timescale t from fast initial rotation ($\omega \gg n$) to a rotation period commensurate with the orbital period ($\omega \sim n$) can be approximated by

$$\tau = \frac{2Ca^6 Q\omega_{\mathrm{initial}}}{3k_2 GM_{\odot}^2 R^5} \qquad [58]$$

As a result of the strong distance dependence ($\tau \propto a^6$), Mercury and Venus can despin within the age of the solar system, whereas Mars cannot (Goldreich and Soter, 1966; see also **Sections 10.04.3, 10.04.4**, and **10.04.5**). However, the quality factor Q of the planets is not well known, nor is its frequency (or rotation rate) dependence well understood (see, e.g., Anderson, 1989; Lambeck, 1980; Munk and MacDonald, 1960). Often, it is assumed that Q is constant for large frequencies (for fast rotation, the main tidal frequency $\sigma \approx 2\omega$) as in seismology (e.g., Dahlen and Tromp, 1998) and inversely proportional to frequency for small frequencies (the main tidal frequency can be small for $\omega \sim n$). The latter case corresponds to a constant phase lag Δt, and dissipation then vanishes when the tidal frequency goes to zero. Zero tidal frequency implies that the tidal bulge is static with respect to the planet, and $\delta = 0$. During the rotation history of a planet, the torque and the despinning rate can change through the frequency dependence of Q and as a result of internal planetary changes that can affect C and k_2.

For slowly rotating planets, the sign of the effect of the solar tidal torque on the rotation rate depends on the obliquity, with two regions separated by an equilibrium point for the rotation rate given by

$$\omega_e = \left(\frac{2\cos\varepsilon}{1+\cos^2\varepsilon}\right)n + \mathcal{O}(e^2) \qquad [59]$$

when dissipation is assumed to be linear with frequency (or the time lag constant) (Correia and Laskar, 2003a). When $\omega > \omega_e$, the rotation rate of the planet decreases, whereas it increases for $\omega < \omega_e$. For an initially rapidly rotating planet, the rotation rate tends to decelerate to this equilibrium rotation rate.

Besides changing the rotation rate of a planet, the tidal torque also changes the obliquity as the tidal bulge is carried outside the orbital plane by rotation during the time lag for obliquities different from 0° and 180°. When the time lag is assumed to be constant and the eccentricity is negligibly small, $\varepsilon = 0°$ and $\varepsilon = 180°$ are unstable equilibrium points and $\varepsilon = \arccos(2n/\omega) = \varepsilon_0$ is a stable equilibrium for fast rotation ($\omega > 2n$) (Goldreich and Peale, 1970; Néron de Surgy and Laskar, 1997). Obliquity increases when $\varepsilon < \varepsilon_0$ and decreases otherwise. The stable equilibrium is between 0° and 90° for prograde rotation and tends to zero obliquity for a decreasing rotation rate. Tidal torques can therefore not overturn an initially prograde planet to a retrograde rotation. For slow rotation ($0 < \omega < 2n$), $\varepsilon = 0°$ is a stable equilibrium point and $\varepsilon = 180°$ an unstable point. Because tidal friction tends to drive the rotation rate ω to the slow rotation regime (eqn [59]), obliquity is eventually driven to 0°, and the rotation rate tends to the mean motion (1:1 resonance). For a non-circular orbit, the final tidally evolved state is commensurate but not synchronous with the mean orbital motion. This will be discussed further in **Section 10.04.5** on Mercury.

10.04.2.4.2 Thermal atmospheric tides

A conceptually related dissipation effect on planetary rotation is due to thermal tides in the atmosphere, first shown by Lord Kelvin (Thompson, 1882). The Sun heats the atmosphere and the surface most at the subsolar point and causes pressure changes and mass motion, or thermal tides, in the atmosphere from the subsolar point to the less heated zones. The main thermal tides of the Earth are semidiurnal and the atmospheric 'bulge' is almost perpendicular to the Earth–Sun direction (Chapman and Lindzen, 1970). When one takes into account a not too large response times, it follows that the solar torque on the bulge, for rapidly rotating planets ($\omega > 2n$), has an effect on the rotation rate of the planet opposite to that of the solar torque on the gravitational tidal bulge: it accelerates the rotation rate (e.g., Dobrovolskis and Ingersoll, 1980; Gold and Soter, 1969; Munk and MacDonald, 1960). Because the torque acts on the atmospheric bulge, it is assumed here that the resulting change in angular momentum of the atmosphere is transmitted to the solid body to maintain a rotation of the atmosphere not too dissimilar from that of the solid body.

The pressure perturbation in the atmosphere is inversely proportional to the tidal frequency (Chapman and Lindzen, 1970). Therefore, the effect of thermal tides on the rotation can be neglected for fast rotation ($\omega \gg n$), and a deceleration from an initial fast rotation to a rotation commensurate with the orbital rotation is dominated by the gravitational solid-body tides. Evidently, torques on the atmospheric thermal tides are mainly important for planets with a dense atmosphere, and dissipation models have mainly been developed for the Earth and Venus (e.g., Chapman and Lindzen, 1970; Dobrovolskis and Ingersoll, 1980).

For $\omega > 2n$, the thermal tides tend to bring the obliquity to either one of the two stable equilibrium points, at $\varepsilon = 0°$ and at $\varepsilon = 180°$ for zero eccentricity, depending on the obliquity being smaller or larger than an unstable equilibrium obliquity value. This critical obliquity depends on w and goes to zero for $\omega \gg n$ and to 180° for $\omega \to 2n$ (Correia and Laskar, 2003a). For initial obliquities larger than the critical obliquity but smaller than 90°, thermal tides can therefore tend to increase the obliquity and tend to overturn a planet that initially rotates rapidly in the prograde sense (Dobrovolskis, 1980; Lago and Cazenave, 1979; Yoder, 1995).

10.04.2.4.3 Core–mantle friction

Friction torques associated with differential rotation between a liquid core and a solid mantle can also change the rotation of a planet on long timescales. In our analysis of periodic rotation variations, the friction torques were neglected in a lowest-order approximation as their effects are expected to be small on short timescales. Nevertheless, friction torques have an observable effect on the nutations of the Earth (e.g., Buffett, 1992; Buffett and Christensen, 2007; Buffett et al., 2002; Deleplace and Cardin, 2006; Dumberry and Koot, 2012; Mathews and Guo, 2005; Mathews et al., 2002). For secular changes in the rotation, even small dissipation effects can eventually have important effects. Two different friction torques must be considered: the viscous torque due to viscosity of the core liquid and the electromagnetic torque between the core and the mantle resulting from the electric currents in the conducting bottom mantle layer (see also **Section 10.04.5.3**). Both torques are linearly proportional to the difference of the rotation vectors of the mantle and core (e.g., Rochester, 1976). For the electromagnetic torque, this can be understood by noting that the magnetic field involved in the coupling is induced by the relative velocity between the mantle and the core. Therefore, usually a single friction torque is considered with parameters that include the effects of both the viscous coupling and the electromagnetic coupling.

In response to precession, an additional flow of uniform vorticity is set up in the core (Hough, 1895, Poincaré, 1910). Without core viscosity, the vorticity of this flow is in the direction of $\vec{\omega} \times \left(\vec{\Omega}_P \times \vec{\omega}\right)$, where $\vec{\Omega}_P$ is the angular velocity of precession oriented normal to the orbital plane (e.g., Stewartson and Roberts, 1963). The vorticity is thus in the $\left(\vec{\omega}, \vec{\Omega}_P\right)$-plane, and the core rotation vector is coplanar with $\vec{\omega}$ and $\vec{\Omega}_P$. Since the additional flow is steady in a precessing frame (the core coprecesses with the mantle), the change in the additional core angular momentum is in the direction of $\vec{\Omega}_P \times \left[\vec{\omega} \times \left(\vec{\Omega}_P \times \vec{\omega}\right)\right]$ (Pais et al., 1999). By using the conservation of angular momentum, it follows that the change in the angular momentum $I\vec{\omega} \approx C\vec{\omega}$ along the rotation axis of the mantle is in the opposite direction. Because that vector is perpendicular to $\vec{\omega}$ and $\vec{\Omega}_P$, there is no corresponding change in $H = C\omega$ nor in $X = C\omega\cos\varepsilon$, so neither the obliquity nor the rotation rate changes.

Core viscosity tends to decrease the relative velocity between the mantle and the core and to make the core precession to lag behind the mantle precession. The additional flow vorticity thus has a component in the direction perpendicular to $\left(\vec{\omega}, \vec{\Omega}_{\mathrm{P}}\right)$ (Busse, 1968; Roberts and Stewartson, 1965; Rochester, 1976; Stewartson and Roberts, 1963), and the change in angular momentum $C\vec{\omega}$ has an additional component in the direction of $\vec{\Omega}_{\mathrm{P}} \times \left(\vec{\omega} \times \vec{\Omega}_{\mathrm{P}}\right)$, which is perpendicular to $\vec{\Omega}_{\mathrm{P}}$ but not to $\vec{\omega}$. As a result, H changes in time and $X = H\cos\varepsilon$ remains constant. When the rotation rate changes because of core–mantle friction, the obliquity ε must also change to keep the component of the spin angular momentum normal to the orbital plane constant:

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = \cot\varepsilon \frac{1}{\omega}\frac{\mathrm{d}\omega}{\mathrm{d}t} \qquad [60]$$

Since the dissipated energy must come from the rotation, core–mantle friction torques cause the rotation to slow down and the obliquity to evolve to either 0° or 180°, depending on the initial obliquity being smaller or larger than 90° (Correia and Laskar, 2003a; Goldreich and Peale, 1970; Néron de Surgy and Laskar, 1997; Pais et al., 1999; Rochester, 1976; Yoder, 1995).

A main problem of core–mantle friction models is that the viscosity of the cores of terrestrial planets is poorly known (Anderson, 1989). Moreover, a turbulent boundary layer close to the CMB can develop (e.g., Yoder, 1995). Core–mantle friction becomes more important for increasing core differential rotation with respect to mantle during precession, and at the same time, turbulence is more likely then. If the FCN period is much smaller than the precession period (such as for rapidly rotating planets as the Earth and Mars at present), the core efficiently tracks the precession of the mantle, and the effect of core–mantle friction on the long-term changes of the rotation of the planet is weak (for the effect on the Earth, see, e.g., Néron de Surgy and Laskar, 1997; Pais et al., 1999). For slowly rotating planets,

$$\frac{P_{\mathrm{FCN}}}{P_{\mathrm{Pression}}} \propto \left(\frac{n}{\omega}\right)^2 \qquad [61]$$

(see eqns [5] and [48]), and the differential rotation between the core and the mantle can become larger. Therefore, core–mantle friction can become the dominant cause of secular rotation variations if rotation is slow enough.

10.04.2.4.4 Climate friction

It is well known that obliquity variations induce climate changes. However, the climate can also change the obliquity of a planet by altering the mass distribution in the planet (Rubincam, 1990). In particular, during glacial cycles, mass is transported between the fluid layers (atmosphere and oceans) and the polar caps. This changing surface load also induces a partially compensating internal mass redistribution. Together, they lead to polar motion and small variations in the rotation rate on the timescale of the glacial cycles (Munk and MacDonald, 1960). Climate cycles also act to modify the polar flattening of the planet, and therefore, the precession rate will vary. For the Earth, the oblateness can change by as much as 1% (Thomson, 1990).

Due to the fact that the mantle is viscous, the changes in the equatorial bulge will lag behind the obliquity changes and will result in a long-term effect on the obliquity. This mechanism was first described by Rubincam (1990) and called 'climate friction.' Despite many unknowns in the climate and viscous response to obliquity oscillations, most recent analyses suggest that the climate friction effect on the long-term obliquity variations of the Earth and Mars is small compared to the planetary effect on obliquity, with obliquity changes below 1° during the age of the solar system for Mars and Earth obliquity changes smaller than 0.01° per My during the major recent Pliocene–Pleistocene glaciations (Levrard and Laskar, 2003; Spada and Alfonsi, 1998 but see also Williams et al., 1998).

10.04.3 Mars

10.04.3.1 Current Rotation State

Christiaan Huygens was the first to observe that the rotation period of Mars is similar to that of the Earth. On 1 December 1659, he noted (in Latin) in his journal that "the rotation period of Mars, like that of the Earth, seems to be in a period of twenty-four hours" (see Kieffer et al., 1992). The period was confirmed in 1666 by Giovanni Domenico Cassini, who determined a value of 24 h and 40 min. Herschel (1774) estimated the obliquity of Mars to be about 30°. The rotation period determined from Doppler and range measurements to the Mars Pathfinder lander and the Viking landers is 88642.66302 ± 0.00002 s, or 24 h 37 min 22.66302 ± 0.00002 s, and the obliquity $\varepsilon = 25.189417 \pm 0.000033°$ (Folkner et al., 1997a). The inclusion of Doppler and range tracking data from the orbital missions Mars Global Surveyor (MGS), Odyssey, and Mars Reconnaissance Orbiter (MRO) slightly improved the estimates of the rotation period 24 h 37 min 22.662993 ± 0.000003 s and obliquity $\varepsilon = 25.1893792424 \pm 0.000010°$ (Konopliv et al., 2011). The spin period here is relative to the node of the Mars mean orbit and Mars mean equator of date.

10.04.3.2 Long-Term Spin Evolution

Because Mars is far from the Sun and has no large moons, tidal braking has only had a small influence on its rotation (Goldreich and Soter, 1966). As the effects of thermal tides in the tenuous Martian atmosphere and of core–mantle friction for a rapidly rotating planet are even smaller (see **Section 10.04.2.4**), the rotation speed of Mars can be considered to be primordial (Laskar et al., 2004). The dissipative effects on obliquity variations are equally small, and planetary perturbations on the orbital motion of Mars dominate the changes in obliquity. The obliquity variations depend sensitively on the precession rate (see **Section 10.04.2.3**), which is well known with a precision below 1%:

$$\dot{\psi} = (-7594 \pm 10) \times 10^{-3} \text{arcsec year}^{-1} \qquad [62]$$

(Konopliv et al., 2011). This estimate has been determined from the Doppler and range data to the Mars Pathfinder lander and the Viking landers and from radio-tracking data of MGS,

Odyssey, and MRO and is close to the value determined on the basis of lander data only (Folkner et al., 1997a).

Because the current precession frequency of Mars is close to some major secular frequencies of the orbital precession, Mars' obliquity performs large chaotic variations (Laskar and Robutel, 1993; Touma and Wisdom, 1993; Ward, 1974). Therefore, the evolution of obliquity cannot be determined over long timescales, and the present obliquity cannot be considered as being primordial. In a statistical study of the orbital and rotational evolution of Mars, Laskar et al. (2004) demonstrated that the current accurately determined precession rate allows constructing reliable obliquity evolutions over 10–20 My. Sufficiently accurate initial conditions such as the current precession rate are needed because the motion of Mars' spin axis is chaotic. These authors also calculated the mean and most probable obliquity values over 4 Gyr: Mars seems to have a much lower obliquity now than its mean value of about 37.62° and its most probable value of about 41.80°. Obliquity and precession are important constraints on the solar insolation and hence on the climate history and the evolution of the polar caps of Mars. In particular, the present 'low' obliquity suggests that the present climate is not representative. We refer to Laskar et al. (2004) and references therein for an extensive study of the astronomical forcing of the Martian paleoclimate.

By using eqn [5], and correcting for the relativistic geodetic precession of 6.7 mas year^{-1} (Reasenberg and King, 1979) and for the precession of -0.2 mas year^{-1} caused by the other planets, the polar moment of inertia C of Mars can be derived from the precession rate as $C=(0.3664\pm0.0005)\ MR^2$ (Konopliv et al., 2011). All other quantities in eqn [5] for the precession rate, including the coefficient J_2 of the gravitational field, are well known. The moment of inertia measures the radial mass concentration and is an important and the best determined constraint on the interior structure of Mars (see **Chapter 10.02**).

10.04.3.3 Nutation

The observation of nutations of Mars can provide insight into the interior of Mars. If Mars were to behave rigidly, nutations would only give a constraint on the moment of inertia, in addition to the constraint from precession. However, the nonrigid behavior contains important geophysical information, in particular on the core. Because the nonrigid contributions to the nutation amplitudes are currently below the limit of detectability, theoretical nutation models are used, and needed, in the analyses of radio science data (Konopliv et al., 2006; Yoder and Standish, 1997). Improved radio science experiments could allow distinguishing the nonrigid contributions and so offer a direct way to study properties of the deep interior of Mars. A small network of a few fixed landers with dual-band coherent radio transponders would be particularly suited. In 2016, NASA will launch a single geophysical lander, InSight, which will enable measuring Mars' nutations. In this section, we review the current theoretical knowledge on the nutation of Mars and discuss the geophysical potential of nutation observations.

10.04.3.3.1 Rigid nutations

The celestial mechanical problem of determining the tidal potential is most conveniently separated from the geophysical problem of calculating the rotational response to forcing by first deriving the nutations under the assumption that Mars responds as a rigid body to the gravitational (tidal) torques. Rigid nutation series for the angular momentum axis can be easily derived from eqn [1] by explicitly expressing the three components in terms of an inertial reference frame. After rewriting the torque components in terms of another nonrotating reference frame with first axis toward the ascending node of Mars' orbital plane on Mars' equator of date and the third axis in the direction of the angular momentum axis, and approximating the magnitude of the angular momentum vector by $C\Omega$ since the equatorial components of the angular momentum are very small compared to the polar component, we obtain

$$\frac{\mathrm{d}\Delta\psi}{\mathrm{d}t}=\frac{1}{C\Omega}\frac{L_1}{\sin\varepsilon} \qquad [63]$$

$$\frac{\mathrm{d}\Delta\varepsilon}{\mathrm{d}t}=\frac{1}{C\Omega}L_2 \qquad [64]$$

where $\Delta\psi$ and $\Delta\varepsilon$ are the periodic nutations in longitude and obliquity and (L_1, L_2) are the equatorial components of the torque with the first component in the direction of the ascending node of Mars' orbital plane on Mars' equator of date (Moritz and Mueller, 1987; Roosbeek, 1999). Equivalent expressions in a Hamiltonian theory are given in Bouquillon and Souchay (1999). For each interacting body of the solar system sufficiently far from Mars such that the torque can be approximated by its degree-two terms, we have

$$\frac{\mathrm{d}\Delta\psi}{\mathrm{d}t}=\frac{1}{\Omega}\frac{1}{\sin\varepsilon}\frac{GM}{r^3}\frac{\left(C-\overline{A}\right)}{C}\frac{3}{2}\sin2\delta\sin\alpha \qquad [65]$$

$$\frac{\mathrm{d}\Delta\varepsilon}{\mathrm{d}t}=-\frac{1}{\Omega}\frac{GM}{r^3}\frac{\left(C-\overline{A}\right)}{C}\frac{3}{2}\sin2\delta\cos\alpha \qquad [66]$$

where M is the mass of the perturbing body, r the distance to the perturbing body, and δ and α the declination and right ascension of the perturbing body, respectively. Triaxial effects of Mars are neglected here and $\overline{A}=(A+B)/2$. The equatorial torque components and the nutations can then be calculated from the ephemerides of the planets and moons (e.g., Borderies, 1980; Hilton, 1991; Reasenberg and King, 1979). The most recent and accurate nutation series (Bouquillon and Souchay, 1999; Roosbeek, 1999) are accurate up to 0.1 mas and are based on the VSOP87 ephemerides for the planets (Bretagnon and Francou, 1988), which are valid for a few 1000 years around J2000.0 and include orbit perturbations due to the other planets. For Phobos and Deimos, the ESAPHO and ESADE ephemerides were used (Chapront-Touzé, 1990).

At a given tidal frequency σ_j', nutation can be expressed either as a sum of a sine and a cosine term in both longitude and obliquity or as a sum of two circular nutations, one prograde and the other retrograde (as compared to the rotation of the planet):

$$\begin{aligned}\Delta\varepsilon+\mathrm{i}\sin\varepsilon_0\Delta\psi&=\sum_j\left[A_j\mathrm{e}^{\mathrm{i}\left(\sigma_j't+f_j\right)}-A_{-j}\mathrm{e}^{-\mathrm{i}\left(\sigma_j't+f_j\right)}\right]\\&=\sum_j\left[\left(A_j-A_{-j}\right)\cos\left(\sigma_j't+f_j\right)+\mathrm{i}\left(A_j+A_{-j}\right)\sin\left(\sigma_j't+f_j\right)\right],\end{aligned} \qquad [67]$$

where ε_0 is the obliquity at J2000.0 (see, e.g., Defraigne et al., 1995).

The largest nutation of Mars considered as a rigid body is the prograde semiannual nutation with an amplitude of about 500 mas (Bouquillon and Souchay, 1999; Roosbeek, 1999). Semiannual here refers to a period of half a Martian year. For the ideal case of a single biaxial planet in circular orbit about the Sun, the semiannual nutation would be the only nutation. In that case, the torque on the planet is equal and maximal at summer and winter solstices. Other nutations are due to deviations with respect to this ideal case. The amplitude of the semiannual nutation is almost equal to that of the Earth. Three factors in the nutation amplitude contribute to this near equality. First, the torque depends on the third power of the distance of the planet to the Sun, which is about 1.5 times larger for Mars than for the Earth. Secondly, the nutation amplitude is proportional to the period (see eqn [19]), which is about 1.88 times longer for Mars than for the Earth. These two factors can be combined by Kepler's third law into a nutation amplitude inversely proportional to the period. Thirdly, the dynamic flattening $H_D=(C-\overline{A})/C$ of Mars is about 5/3 of the flattening of the Earth.

The amplitudes of the four main retrograde nutations are about 137, 18, 5, and 1 mas for periods of 1, 1/2, 1/3, and 1/4 year, respectively (Roosbeek, 1999; values have been rescaled to use the most recent precession estimate of Konopliv et al., 2011). The prograde nutations at 1, 1/2, 1/3, and 1/4 year have amplitudes of 102, 499, 108, and 18 mas. All these nutations are due to the gravitational torque of the Sun. The nutations due to Mars' moons have amplitudes of about 10 mas for Phobos and 5 mas for Deimos and periods of 2.26 years and 54.754 years, respectively (Bouquillon and Souchay, 1999; Roosbeek, 1999). The largest nutations caused by the direct gravitational torque from planets are on the order of 0.1 mas and are due to Jupiter and the Earth.

Differences between the nutations for the angular momentum, the rotation axis, and the figure axis are at most on the order of 0.1 mas and can be neglected (Bouquillon and Souchay, 1999; Roosbeek, 1999). The largest uncertainty on the rigid nutation amplitudes is due to the observational error on the precession rate. Both precession and nutation depend linearly on the dynamic flattening $H_D=(C-\overline{A})/C$ (Roosbeek, 1999). Any relative error on the precession rate translates into the same relative error on H_D and on the nutations. With the newest and most accurate precession rate of Konopliv et al. (2011), the value used in the nutation amplitudes mentioned earlier, the relative uncertainty of 0.13% on the precession rate implies an uncertainty on the main nutation of 0.6 mas.

10.04.3.3.2 Nonrigid nutations

A nonrigid nutation series is derived from a rigid nutation series by means of a transfer function $T(\sigma)$, which describes the ratio of the nonrigid nutation amplitude to the rigid nutation amplitude for frequency σ (Dehant et al., 2000a). We will use a transfer function for a planet with a liquid core because Mars most likely has a – at least partially – liquid core as indicated by the fact that the Love number $k_2=0.164\pm0.009$ (Konopliv et al., 2011) is clearly larger than the value of about 0.07 for an entirely solid Mars (Rivoldini et al., 2011; Van Hoolst et al., 2003; Yoder and Standish, 1997). Thermal evolution models also point to an at least partially liquid core (Schubert and Spohn, 1990; Stevenson et al., 1983). For a planet with a solid mantle and a liquid core, the transfer function can be derived from Liouville equations [23]–[25] and [32]–[34] as

$$T(\sigma)=\frac{(\sigma-\sigma_E)\left(A\sigma'-A^m\sigma'_{FCN}\right)}{A^m(\sigma-\sigma_{CW})(\sigma-\sigma_{FCN})} \qquad [68]$$

which shows that nutations are resonant at frequencies equal to the CW frequency and the FCN frequency. For a planet containing also a solid inner core, resonances with the FICN and the ICW must additionally be considered. The FCN is particularly important because it has a long retrograde period in an inertial reference frame between about 200 and 300 days, depending on the interior structure, close to the periods of the main nutations (see eqn [48], Hilton, 1992; Van Hoolst et al., 2000b; Yoder and Standish, 1997). Therefore, the FCN can have a large effect on nutation and the amplifications, and amplitudes, of the retrograde nutations depend sensitively on the model (Dehant et al., 2000a,b; Le Maistre et al., 2012; Molodensky et al., 2009; see **Figure 2**). For the Earth, relatively large resonant nutation amplifications have been detected and have been used to demonstrate that the flattening of the core is larger than that of a hydrostatic Earth based on the PREM model (Gwinn et al., 1986; Mathews et al., 1991, 2002).

The amplification of the largest nutation with prograde semiannual period depends much less on the FCN period since it is much further from the retrograde FCN period and thus from the resonance peak. Nevertheless, as it is the largest nutation, its amplification with respect to the rigid nutation term is still about 5 mas, for any realistic model of Mars' interior. Future determinations of the amplitudes of nutation terms with a radio science experiment are expected to have a precision on the order of a mas (Dehant et al., 2000a;

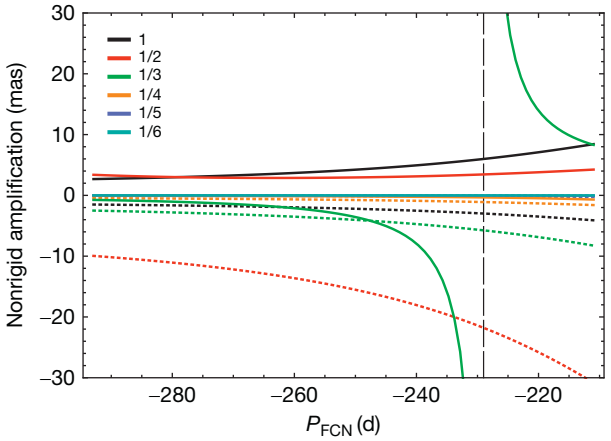


Figure 2 Amplification of prograde (dashed lines) and retrograde (solid lines) nutations with respect to rigid nutation amplitudes for the nutations with periods equal to 1 Martian year (1), half a Martian year (1/2), and up to one-sixth of a Martian year (1/6) as a function of the FCN period for a selected set of basic two-layer models of the interior structure of Mars satisfying the observed mass and moment of inertia.

Le Maistre et al., 2012). Only for the prograde semiannual nutation is the amplification always above the observational precision and larger than the uncertainty on the rigid amplitude, irrespective of the interior structure of Mars. Therefore, this nutation is ideally suited for further independent evidence of the liquidity of the core.

Besides establishing the liquidity of the Martian core, the nutations can be used to constrain other properties of the interior of Mars, by making use of the resonances of the main nutations with rotational normal modes whose frequencies depend sensitively on Mars' interior. The FCN period depends mainly on the CMB polar flattening α^f because the restoring force for this normal mode is proportional to that flattening and on the moment of inertia of the core. It therefore decreases in absolute values with increasing core radius and is a good indicator of core density (Van Hoolst et al., 2000a). Because the FCN eigenfunction represents essentially rotational motion, the influence of uncertainties on rigidity, incompressibility, and mantle inelasticity on the FCN period is small (Hilton, 1992; Le Maistre et al., 2012; Molodensky et al., 2009; Van Hoolst et al., 2000a,b). For realistic models of Mars' interior in hydrostatic equilibrium with an inner core radius of about 1800 km, a mean value derived from the observed k_2 Love number (Rivoldini et al., 2011), the retrograde FCN period is expected to be between about 220 and 280 days (Le Maistre et al., 2012) so that a resonant amplification of the terannual retrograde nutation is possible.

The closer the FCN period is to one of Mars' main nutations, the more accurately the FCN period and thus the core radius and density of Mars can be determined. After 50 weeks of Doppler tracking observations, with a precision of 1 mm s^{-1}, between four landers and an orbiter, and the orbiter and the Earth, the FCN period is expected to be determined with a 10-day uncertainty (Yseboodt et al., 2003) for an FCN well separated from the terannual period. For one nearly equatorial lander, the FCN period is estimated with a similar precision after 550 days of mission (Le Maistre et al., 2012). For an FCN period equal to −230 days, close to the terannual period of 228.9 days, a precision of 1 day is reached after 300 days of observations. After 1 year, a precision of 30% on the core moment of inertia and of 5% on the CMB flattening can be obtained (Le Maistre et al., 2012).

Mantle convection and the presence of a solid inner core could complicate the interpretation of the nutation amplitudes. Mantle convection can deform the CMB and change the core flattening. In particular, the Tharsis region, which is the dominant topography on Mars, is partially associated with a hot upwelling plume in the mantle (Harder and Christensen, 1996; Kiefer et al., 1996; Roberts and Zhong, 2004). As Tharsis is situated in an equatorial area, it increases the polar flattening of the core. Steady state convection models that account for the geoid high at Tharsis show that the FCN frequency can be decreased down to about 100 days, in particular when no lower mantle perovskite layer is present (Defraigne et al., 2001). The FCN resonance then shifts to smaller rigid nutation terms (see **Figure 2**). Yoder et al. (2003) showed that at least the outer part of the core is liquid, but it is not known with certainty whether there would be a solid inner core, nor how large it would be. The main uncertainties are the temperature profile and the concentration of light elements in the core, especially sulfur, which has a large effect on the melting temperature of the core assemblage (Schubert and Spohn, 1990; Spohn, 1991). Using SNC meteorites, Dreibus and Wänke (1985) showed that the percentage of sulfur in the core is about 14 wt.%. Such a high sulfur content tends to keep the core entirely liquid (Breuer and Spohn, 2003; Schubert and Spohn, 1990; Schubert et al., 1992). In a recent analysis of the observed k_2 Love numbers, this large sulfur content was confirmed, under the condition that sulfur is the only light element in the Martian core, and the core was concluded to be entirely liquid for realistic temperature profiles (Rivoldini et al., 2011).

The absence of a magnetic field of a global scale has also been used to argue against the existence of a solid inner core (e.g., Breuer and Spohn, 2003) since a growing inner core is very efficient in driving a dynamo. Nevertheless, core growth could be too small to sustain a dynamo, might be in a regime where the solidifying material has the same composition as the liquid, or could have stopped due to core reheating after an early period of global plate tectonics (Stevenson, 2001). Besides dynamo studies and thermal evolution calculations, observations are needed to decide upon the existence of an inner core. Nutation observations could help in deciding about the existence and size of the Martian inner core. When the outer core has a sulfur concentration below the eutectic value, the inner core influence on the nutations is below 1 mas except if the FICN period happens to be very close to the prograde annual or semiannual nutation (Defraigne et al., 2003; Dehant et al., 2003; Van Hoolst et al., 2000b). The observation of such an amplification would be a strong indication for the existence of an inner core but would require a lucky coincidence. For progressively larger inner core, the amplification due to the FCN resonance of the largest nutation, the prograde semiannual, is almost canceled by the resonance with the FICN, whose period decreases rapidly to zero in the limit of an entirely solid core (Defraigne et al., 2003). Failure to detect amplification of the semiannual nutation with a future sufficiently accurate geodesy experiment could then be interpreted as evidence for a large inner core.

10.04.3.4 Rotation Speed Variations

The atmosphere of Mars consists of 95% carbon dioxide (CO_2) (Owen et al., 1977), which condenses during the winter seasons at high latitudes to form periodic polar CO_2 caps. During spring and summer seasons, the caps sublime and the mass is again redistributed in the atmosphere (James et al., 1992). In this seasonal cycle of mass transfer between the atmosphere and the polar caps, a large fraction of the total mass of the atmosphere condenses at the polar regions. The maximum mass of the north and south seasonal caps is $3.0 \pm 1.0 \times 10^{15}$ kg and $5.5 \pm 1.0 \times 10^{15}$ kg, respectively (Karatekin et al., 2006b; Konopliv et al., 2006; Litvak et al., 2006; Prettyman et al., 2009), or about one-quarter of the total mass of the atmosphere of about 2.5×10^{16} kg (Lodders and Fegley, 1998). Because of the conservation of angular momentum, this large mass redistribution by seasonal sublimation and condensation of CO_2 and the (associated) winds can

induce seasonal changes in the rotation of Mars. The more the mass is concentrated in the polar regions, as solid CO_2 ice in the polar caps, the smaller the moment of inertia C about the polar axis is and the faster Mars rotates. When the polar caps sublime, the CO_2 is distributed more uniformly over the Martian surface and Mars rotates slower. Qualitatively, the effect on rotation scales as the ratio of the net redistributed mass to the planet's total mass (Chao and Rubincam, 1990). This ratio is about 1.3×10^{-8}, at least one to two orders of magnitude larger than any mass redistribution on Earth, suggesting large rotation rate variations.

The redistribution of matter is mainly in the north/south direction. If the growth of one polar cap would be symmetrical with respect to the retreat of the other polar cap, there would be no net change in the moment of inertia of Mars, and only winds in the atmosphere could change the rotation rate. In reality, the caps do not grow symmetrically. Since the orbit of Mars is quite eccentric (eccentricity equal to 0.0934), the temperatures at the two poles differ for a given season, and the seasons do not have the same length on both hemispheres. For example, winter on the southern hemisphere lasts almost a month longer than on the northern hemisphere (Kieffer et al., 1992). Differences in polar topography, thermal inertia, and albedo further contribute to asymmetry between the increase and the decrease of the two polar caps.

Seasonal changes in the rotation speed of Mars have first been detected from radio-tracking data to the Viking landers (Folkner et al., 1997a; Reasenberg et al., 1980; Yoder and Standish, 1997). Although the rotation rate variations can best be determined from a lander, more recent solutions are based on MGS, Odyssey, and MRO tracking data, which are both qualitatively and quantitatively better than the older Viking data. Solutions based on the orbiter tracking data are consistent with the lander data within the uncertainties (Konopliv et al., 2006). Using both lander and orbiter data, Konopliv et al. (2011) determined amplitudes of 456 ± 23 mas, 154 ± 21 mas, 24 ± 19 mas, and 10 ± 16 mas for the rotation angle variations at 1-, 1/2-, 1/3-, and 1/4-year period, respectively. In these studies, a theoretical nutation model is included. If it were omitted, it would substantially change the estimates of the seasonal rotation variations (see, e.g., Yoder et al., 2003).

Instead of rotation angle variations, relative changes in the rotation rate $\delta\omega_3/\Omega = m_3$, changes in the length of the day, $\mathrm{LOD}(t) = 2\pi/\omega_3(t)$, or variations in the time needed for the planet to rotate over the angle $\phi(t)$ at constant angular velocity Ω are often used to describe rotation variations. The latter variations are denoted as δUT by analogy with LOD variations for the Earth. We have the following useful relations between the rotation variations:

$$\frac{\Delta\mathrm{LOD}(t)}{\mathrm{LOD}} = -m_3(t), \qquad [69]$$

$$\delta\phi(t) = \Omega \int_{t_0}^{t} m_3(t')\mathrm{d}t' \qquad [70]$$

$$\delta\mathrm{UT}(t) = \int_{t_0}^{t} m_3(t')\mathrm{d}t' \qquad [71]$$

$$= \frac{1}{\Omega}\delta\phi(t) \qquad [72]$$

$$= -\frac{\Omega}{2\pi}\int_{t_0}^{t} \Delta\mathrm{LOD}(t')\mathrm{d}t' \qquad [73]$$

For a periodic variation with frequency ω, the amplitudes A_{LOD}, A_{UT}, A_{m_3}, and A_ϕ of the variations in LOD, UT, m_3, and ϕ, respectively, are related by

$$A_{m_3} = \omega A_{\mathrm{UT}} = \frac{\Omega}{2\pi} A_{\mathrm{LOD}} = \frac{\omega}{\Omega} A_\phi \qquad [74]$$

As an example, the observed amplitude $A_\phi = 456$ mas of the annual rotation variations (Konopliv et al., 2006) corresponds to amplitudes $A_{\mathrm{UT}} = 31.9$ ms, $A_{m_3} = 3.37 \times 10^{-9}$, and $A_{\mathrm{LOD}} = 0.299$ ms. For the semiannual amplitude, we have $A_\phi = 154$ mas, $A_{\mathrm{UT}} = 10.6$ ms, $A_{m_3} = 2.23 \times 10^{-9}$, and $A_{\mathrm{LOD}} = 0.198$ ms. Hereafter, only rotation variations in mas will be quoted.

With modern radio tracking of a lander over at least one Martian year, the accuracies on the rotation variations could be increased. Because the older Viking data are an order of magnitude less precise than modern X-band tracking data and the rotation estimates based on them and on Mars Pathfinder data are of comparable accuracy to those based on the modern spacecraft data, an order of magnitude improvement can be expected. Direct two-way range and X-band Doppler measurements between an Earth tracking station and a lander on Mars, or Earth-based radio-tracking observations of an orbiter that is itself tracking one or more landers, would enable the determination of the rotation variations with an accuracy of a few mas (Barriot et al., 2001; Dehant et al., 2009, 2011; Folkner et al., 1997b; Le Maistre et al, 2012; Yseboodt et al., 2003).

Theoretical estimates of rotation rate variations were first derived by Philip (1979), who used Viking data to calculate the atmospheric friction torque associated with zonal winds. This torque transfers angular momentum between the atmosphere and the solid planet and causes rotation rate variations. Although the torque approach gives insight into the mechanisms transferring angular momentum between the solid body and the atmosphere, most studies have used an angular moment approach as developed in **Section 10.04.2** since this has proven to be more reliable (see also Sanchez et al., 2003). Wind and loading data to calculate the excitation functions or the angular momentum functions of **Section 10.04.2** can be obtained from general circulation models (GCMs) of the Martian atmosphere or from physical considerations on mass redistribution within the Martian atmosphere and polar caps (see, e.g., Chao and Rubincam, 1990; Yoder and Standish, 1997). Recent GCMs were used by Van den Acker et al. (2002) and Sanchez et al. (2003). The former study used the Martian Climate Database v3 from the Laboratoire de Météorologie Dynamique (LMD) in Paris, France (Forget et al., 1999; Lewis et al., 1999), while the latter the NASA Ames GCM (Pollack et al., 1990).

The estimated annual and semiannual rotation angle amplitudes are 394 and 192 mas, respectively, in the study of Van den Acker et al. (2002) and correspond reasonably well with the observational values of Konopliv et al. (2006) within their uncertainties. The annual and semiannual amplitudes of

Sanchez et al. (2003) are 291 and 106 mas, respectively. The latter study did not include effects of elastic deformation and of a liquid core. The loading effect of Van den Acker et al. (2002) is in better agreement with the theoretical estimates of Yoder and Standish (1997), who not only used a simpler model for the atmospheric and ice cap loading (see also Chao and Rubincam, 1990) but also included effects of elastic yielding and a liquid core, suggesting that these effects could explain the difference with Sanchez et al. (2003). The deformational effect decreases the rotation variations by about 8% (Defraigne et al., 2000). As the ratio C/C^{m} is only about 1.03 (the value for the standard model A of Sohl and Spohn, 1997; see, e.g., Van Hoolst et al., 2000b), a liquid core increases LOD variations at most by about 3%. However, this effect is to a large extent compensated by the different deformational effect (k_2' in eqn [45]). Deformation tends to decrease rotation variations, and its effect on rotation is a few percent larger for a liquid core than for a solid core. As a result, the effect of the state of the core is very small and on the order of 1% (Defraigne et al., 2000; Karatekin et al., 2006a). The differences between the results of Van den Acker et al. (2002) and Sanchez et al. (2003) can therefore not be explained in terms of the neglect of these effects by Sanchez et al. (2003), but are most likely due to differences in the atmospheric models. This indicates that ΔLOD observations are useful to constrain the GCMs. Note that rotation differences between GCM models having different atmospheric dust contents and dust storms with respect to the standard model for the 'mean' climate of Mars are about one order of magnitude larger than the expected accuracy of future ΔLOD measurements (Defraigne et al., 2000; Van den Acker et al., 2002).

The separate effects of wind, atmospheric loading, and ice cap loading can be studied with the angular momentum approach. The winds have the smallest effect on the rotation rate of Mars in contrast with the Earth where the winds are the dominant source for seasonal LOD variations (Cazenave and Balmino, 1981; Defraigne et al., 2000; Gross et al., 2004; Van den Acker et al., 2002; Yoder and Standish, 1997). From the data of a GCM of the Martian atmosphere developed at the LMD in Paris, France (Forget et al., 1999), wind is found to induce rotation angle variations with an amplitude of 14 mas for the annual period and 76 mas for the semiannual period (Van den Acker et al., 2002). The contributions from atmosphere and ice cap loading are almost out of phase, with the atmosphere having an effect almost three times as large as that of the ice caps (Defraigne et al., 2000; Van den Acker et al., 2002). The effect of the ice caps is smaller than that of the atmospheric loading because mass at the polar ice caps has only a small contribution to the polar moment of inertia C, whereas atmospheric masses can be a Mars radius further from the polar axis and cause a larger change in the polar moment of inertia.

By neglecting the wind contribution and the small effects on rotation variations due to the elastic yielding of Mars and to the presence of a liquid core, LOD variations can most simply be obtained from the conservation of angular momentum $C\omega_3$. We here assume principal axis rotation, which is correct up to the first order in the rotation variations (see Section 10.04.2). The variation in the relative rotation rate is then proportional to the relative change in the largest moment of inertia C:

$$\frac{\Delta\omega_3}{\Omega} = -\frac{\Delta C}{C} \qquad [75]$$

From the definition of the gravitational coefficient $J_2 = [C - (A+B)/2]/MR^2$ and relation (Dobrovolskis, 1980) between rotation rate variations and LOD variations, we have

$$\frac{\Delta \mathrm{LOD}}{\mathrm{LOD}} = \frac{2}{3} MR^2 \frac{\Delta J_2}{C}. \qquad [76]$$

LOD variations can thus directly be determined from estimates of ΔJ_2. The contribution of the atmosphere to ΔJ_2 is about an order of magnitude smaller than that of the ice caps (Chao and Rubincam, 1990; Smith et al., 1999). LOD variations can then be determined from polar ice cap data only. In the approach with angular momentum functions described earlier, the atmosphere pressure is more important than the polar caps. This seemingly contradictory finding is due to fact that the difference between the terms in the ΔJ_2 approach indicates a difference in J_2, not in the rotation variations. To relate ΔJ_2 to rotation variations, mass conservation is assumed, which is not satisfied for the atmosphere or the ice caps separately. By using eqn [76], observational estimates of LOD variations can be obtained from several independent data. First, MGS, Odyssey, and MRO radio-tracking data have been used to estimate variations in the zonal degree-two coefficient of the gravitational field ΔJ_2 (Konopliv et al., 2006, 2011; Smith et al., 2001; Yoder et al., 2003). Secondly, the ice cap contribution to ΔJ_2 can be estimated from cap elevation changes up to 1.52 m measured with the onboard laser altimeter MOLA of MGS (Aharonson et al., 2004; Smith et al., 2001) and from the high-energy neutron detector (HEND, Litvak et al., 2004), the neutron spectrometer (NS) (Prettyman et al., 2009), and thermal imaging (Kieffer and Titus, 2001; Kieffer et al., 2000) instruments on board of Mars Odyssey. Best comparisons are between the LMD GCM and the NS data, but in general, differences are of the order of 10% (Karatekin et al., 2006a; Konopliv et al., 2011).

Besides loading and wind effects, also tides can induce rotation variations. Zonal tides change the moment of inertia C and so, by conservation of angular momentum, the length of day changes. From the change in C for a model with a decoupled liquid core (see also Yoder et al., 1981), amplitudes of the annual and semiannual tidally driven rotation angle variations were estimated by Yoder and Standish (1997) to be about 50 and 30 mas, respectively, or about one-tenth of the amplitudes induced by atmosphere and ice cap loading. Defraigne et al. (2000) numerically integrated the equation of motion, Poisson's equation, and the continuity equation from the center of the planet up to the surface (see Defraigne and Smits, 1999) to determine the rotational response of the planet to the zonal tides accurately determined by Roosbeek (1999). Elastic and inelastic mantles were considered, and the annual and semiannual rotation angle amplitudes are 15 and 8 mas, respectively. The difference between both models is on the order of 1 mas, at the limit of the expected future observational precision.

10.04.3.5 Polar Motion

10.04.3.5.1 Seasonal polar motion

Atmospheric dynamics and the condensation/sublimation cycle of the atmosphere and polar caps not only change the

rotation speed but also cause fluctuations in the orientation of the figure axis with respect to the rotation axis of Mars, partly because the polar caps are not symmetrical with respect to the rotation axis. Up to now, polar motion of Mars has not been measured directly. Theoretical model calculations have shown that the seasonal polar motion amplitudes are at most 20 mas, which is at the present-day detectability limits, but larger than the expected future precision of a few mas (e.g., Barriot et al., 2001; Defraigne et al., 2000; Konopliv et al., 2006; Yoder and Standish, 1997; Yseboodt et al., 2003).

The main atmospheric forcing is at seasonal frequencies, resulting in largest wobble amplitudes of about 10 mas at 1, 1/2, and 1/3 year, the exact values depending on the atmospheric and internal structure models used (Chao and Rubincam, 1990; Defraigne et al., 2000; Sanchez et al., 2004; Van den Acker et al., 2002; Yoder and Standish, 1997). The annual and semiannual wobble amplitudes, respectively, 13.6 and 11.2 mas, observationally determined by Konopliv et al. (2006) agree remarkably well with the theoretical values of 12.0 and 9.8 mas, respectively, of Van den Acker et al. (2002). The fact that the amplitude at 1/3 year is expected to be of the same order of magnitude as the amplitudes at 1 and 1/2 year is most likely due to the resonance with the Chandler wobble with a period of about 200 days (eqn [47], Dehant et al., 2003; Konopliv et al., 2006; Sanchez et al., 2004; Van Hoolst et al., 2000b; Yoder and Standish, 1997; Zharkov and Gudkova, 2005).

If Mars has a solid inner core, also the ICW period could be close to a seasonal forcing frequency, resulting in resonant polar motion amplification (Dehant et al., 2003). The degree of amplification is difficult to estimate and depends, among others, on the Q factor of the ICW, which is unknown. Nevertheless, the resonance strength for the ICW is very small, and a very close approach is needed for noticeable effects (Dehant et al., 2003, 2006).

10.04.3.5.2 Chandler wobble

Besides polar motion at seasonal frequencies, the Martian atmosphere could also excite the Chandler wobble. For the Earth, the Chandler wobble is the largest polar motion signal and is thought to be excited mainly by ocean bottom pressure fluctuations (Gross, 2000). Because Mars has no oceans and the Chandler wobble period for Mars ($P_{CW} \approx 200$ days) is far from the annual and semiannual periods where most excitation power of the atmosphere is expected (Dehant et al., 2006; Sanchez et al., 2004), the CW is not expected to have a large amplitude. From the spectral amplitude in the pressure recordings of the Viking landers near the CW period, Yoder and Standish (1997) estimated a CW amplitude of about 10 mas and possibly as large as 50 mas with ice loading excitation included. Dehant et al. (2006) estimated the power spectral density of the atmospheric excitation close to the CW period from the LMD GCM (Forget et al., 1999) and determined a CW amplitude between 30 and 100 mas for values of the quality factor Q between 80 and 300 (for estimates of the Q factor, see Bills et al., 2005; Smith and Born, 1976; Yoder, 1982). Other excitation sources such as marsquakes are too small (Golombek, 2002; Golombek et al., 1992) to excite polar motion to an observable level (Dehant et al., 2006).

From the observed variations in the order 1 gravity coefficients, and using ice cap models, Konopliv et al. (2006) determined an upper bound on the CW amplitude of about 10 mas. In their analysis based on MGS and Odyssey radio-tracking data, the CW signal in the gravity variations could not be isolated from the 1/3-year forced wobble. Unambiguous detection of the CW and determination of its period could be used to determine interior properties of Mars, although the CW period is mainly constrained by the global dynamic polar and equatorial flattenings of Mars. For an entirely rigid Mars, the CW period would be

$$P_{CW}^{rigid} = \frac{2\pi}{\Omega}\sqrt{\frac{AB}{(C-A)(C-B)}} = 190.8 \pm 0.3 \text{ days} \qquad [77]$$

(Konopliv et al., 2006), much smaller than the corresponding value for the Earth (about 304 days) because of the larger flattening of Mars. The true CW period is most likely between about 190 and 210 days (Dehant et al., 2003; Konopliv et al., 2006; Van Hoolst et al., 2000b; Yoder and Standish, 1997; Zharkov and Gudkova, 2005). The CW only weakly depends on the outer core and on the inner core. The core can have an influence on the order of a few days (Konopliv et al., 2006; Van Hoolst et al., 2000b), and the changes induced by the inner core are smaller than the uncertainty on the global dynamic flattening (Dehant et al., 2003). Large mantle anelasticity (Q-values below 100) could increase the period by up to about 7 days (Van Hoolst et al., 2000b). Uncertainties in the rigidity and compressibility profiles also affect the Love number k_2 and therefore the CW period (see eqn [47]). For large changes of 10% and 21% in the mantle seismic velocities V_P and V_S with respect to a standard model, the CW period changes by about 0.5 days and 5 days, respectively (Van Hoolst et al., 2000a). To derive interior properties of Mars from the CW period, its future observational accuracy must therefore be at least about 1 day.

10.04.4 Venus

10.04.4.1 Current Rotation State

The analyses of radar observations of Venus during its inferior conjunction in fall 1962 showed that Venus rotates retrogradely with a period of about 250 days about an axis nearly perpendicular to its orbit (Carpenter, 1964; Goldstein, 1964). This period is close to the 243.16-day period for the synodic resonance with the orbital motion of the Earth, in which Venus would show exactly the same face to the Earth at each inferior conjunction and would rotate exactly five times as seen from the Sun between consecutive inferior conjunctions, suggesting that Venus might occupy that resonance (Gold and Soter, 1969; Goldreich and Peale, 1966a,b). Further observations with the Goldstone, Arecibo, and Haystack radar systems yielded an improved period of 243.026 ± 0.006 days (Carpenter, 1966, 1970; Dyce et al., 1967; Shapiro et al., 1990), and the analyses of Magellan data have led to an averaged rotation period of 243.0185 ± 0.0001 days based on radar images (Davies et al., 1992) and of 243.0200 ± 0.0006 days based on X-band radio-tracking data (Konopliv et al., 1993, 1999), all three very close to but not exactly equal to the resonance period of 243.16 days. The analyses of Magellan radar data show that the current obliquity of Venus is about 2.6° (or 177.4°) (Davies et al., 1992; Yoder, 1995).

10.04.4.2 Long-Term Spin Evolution

The long rotation period and the retrograde rotation and the almost perpendicular position of Venus with respect to its orbital plane could be primordial. More plausibly however, the rotation of Venus has changed over time. The dissipation of tides in Venus raised by the Sun causes a delay in the appearance of the tidal bulge of Venus, and the solar torque on it tends to decrease the rotation period of Venus and to drive Venus to a 1:1 resonance with its orbital motion on a timescale of 10^8 years to 10^9 years (Correia and Laskar, 2003a; Dobrovolskis and Ingersoll, 1980; Yoder, 1995). Venus' slow rotation can therefore not be taken as an indication for a primordial slow rotation. Instead of being locked in a spin–orbit or spin–synodic resonance, the rotation of Venus is thought to be the result of a balance between solid-body tidal torques, which drive Venus to synchronous rotation, and atmospheric torques, which drive it away (Correia and Laskar, 2001; Correia and Laskar, 2003a, 2003b; Correia et al., 2003; Dobrovolskis, 1980; Dobrovolskis and Ingersoll, 1980; Gold and Soter, 1969; Kundt, 1977; Lago and Cazenave, 1979; McCue and Dormand, 1993; Shen and Zhang, 1989; Yoder, 1995). For the Earth, the thermal tides are too small to balance the tidal gravitational dissipation, but Venus has a much denser atmosphere and, most likely, a smaller tidal dissipation due to the absence of oceans. The acceleration of the Earth's rotation due to the thermal tides is about three times smaller than the deceleration due to the solar gravitational torque and less than 10% of the total solid-body gravitational effect including the Moon (Munk and MacDonald, 1960).

For a fast initial rotation ($\omega \gg n$), both the gravitational tides and friction at the CMB decelerate the rotation of Venus, whereas the atmospheric thermal tides that tend to accelerate Venus are small and their contribution to the rotation rate can be neglected. Once the slow rotation regime ($\omega \sim n$) has been reached, the core–mantle friction torque dominates the obliquity evolution and drives Venus to an orientation with its polar axis perpendicular to its orbital plane for a wide range of values for the dissipative parameters (Correia and Laskar, 2001; Correia et al., 2003; Goldreich and Peale, 1970; Yoder, 1995). For such a final orientation, core–mantle friction does not affect the rotation rate anymore, and thermal atmospheric tides can balance the solid-body tidal braking effect. Starting from a prograde initial rotation, Venus may have evolved according to one of the following two different rotation histories. In the first scenario, Venus' rotation axis flips direction under the influence of atmospheric torques and slows down (Dobrovolskis, 1980; Lago and Cazenave, 1979; McCue and Dormand, 1993; Shen and Zhang, 1989; Yoder, 1995). In the second scenario, Venus slows down while changing its obliquity to 0° and then develops a retrograde rotation (Correia et al., 2003; Kundt, 1977). The current rotation state of Venus therefore places no constraints on the primordial spin.

Correia and Laskar (2001, 2003a) have extensively tested the long-term evolution of the spin of Venus in realistic simulations that include the planetary perturbations. Without planetary perturbations, a final rotation state with $\varepsilon = 180°$ is only possible for sufficiently large initial obliquities and a dominant atmospheric torque (Correia and Laskar, 2001; Dobrovolskis, 1980; Lago and Cazenave, 1979; McCue and Dormand, 1993; Shen and Zhang, 1989; Yoder, 1995). Planetary perturbations can lead to large chaotic changes in the obliquity of Venus as a result of resonance overlap between the precession frequency and combinations of secular frequencies of changes in the orbital plane of Venus (Laskar and Robutel, 1993; Yoder, 1995). Based on eqn [5] for the precession rate and possible values for the polar moment of inertia, the current precession rate of Venus has been estimated to be about 45 arcsec year^{-1}, corresponding to a precession period of about 28 800 years (Yoder, 1997). If Venus initially had an obliquity lower than 90° and a rotation period shorter than 5 days, it could not have escaped the chaotic zone, which lies between 0° and 90° (Correia et al., 2003). As a result, Venus could have reached the final retrograde state with obliquity $\varepsilon = 180°$ starting from any initial obliquity (Correia and Laskar, 2001, 2003a; Correia et al., 2003; Yoder, 1997). Planetary perturbations largely increase the probability of reaching a retrograde final state and make it the most probable final state for small initial rotation periods (< 5 days) or high initial obliquities (Correia and Laskar, 2001, 2003a; Correia et al., 2003).

In the presence of planetary perturbations, capture into the 1:1 spin–orbit resonance becomes unlikely and the obliquity never goes to zero (Correia and Laskar, 2001, 2003a; Correia et al., 2003). Small forced obliquity changes up to a few degrees are possible and are associated with small changes in the rotation rate (Correia et al., 2003; Yoder, 1995). Therefore, with planetary perturbations included, the final states are no longer steady states. The current obliquity of Venus is 2.6° (Davies et al., 1992; Yoder, 1995). Planetary-induced changes in the eccentricity of the orbit of Venus can also lead to long periodic changes in the rotation rate. Because the strengths of the tidal gravitational and thermal atmospheric torques have different dependencies on the radial distance from the Sun, changes in the eccentricity cause slight imbalances in these torques. The resulting rotation velocity changes can be up to a few km year^{-1} at the equator, over timescales of 10^6 years (Bills, 2005b).

10.04.4.3 Wobble, Length-of-Day Variations, and Nutation

A departure of the mean position of the polar principal moment of inertia axis of Venus from the rotation axis has been determined from coherent two-way radio-tracking data of the Pioneer Venus Orbiter (PVO, 1978–92) and Magellan (1990–94). First results suggested a large orientation difference of a few degrees (Bills et al., 1987; Mottinger et al., 1985), much larger than the wobble amplitudes for the Earth and Mars, which are below an arcsec (Munk and MacDonald, 1960, **Section 10.04.3**). Later studies including Magellan data gave estimated values of about 0.5° (Konopliv and Sjogren, 1994; Konopliv et al., 1993, 1999; McNamee et al., 1993), although Nerem et al. (1993) (see also Konopliv et al., 1993), who only used PVO data, obtained an even smaller amplitude of about 0.1°. Konopliv and Sjogren (1994) quote a formal error of 0.05° for the pole position.

Due to the slow rotation of Venus, $J_2 = (4.404 \pm 0.002) \times 10^{-6}$ (Konopliv et al., 1999) is much smaller than for the Earth and not dominated by the contribution from rotation, which is on the order of 10^{-8}. The equatorial flattening is equally small,

with $(B-A)/MR^2 = 4C_{22} = (2.215 \pm 0.003) \times 10^{-6}$. Because of these small flattenings and the long rotation period, the Chandler wobble has a very long period. In the rigid approximation, we have

$$P_{\mathrm{CW}}^{\mathrm{rigid}} = \frac{2\pi}{\Omega}\sqrt{\frac{AB}{(C-A)(C-B)}} \approx 50,000 \text{ years} \qquad [78]$$

(Mottinger et al., 1985). Elasticity tends to increase this period by less than 1% ($k_2 = 0.295 \pm 0.066$, Konopliv and Yoder, 1996), whereas a liquid core would decrease the period by about 10%, depending on the size and density of the core (see eqn [47]). The relative effect of elasticity is much smaller than the almost 50% increase of the CW for Earth due to elasticity (Smith and Dahlen, 1981) because the polar flattening of Venus (or J_2) is mainly due to nonhydrostatic effects and not to rotational flattening as for the Earth. The deformation associated with the Chandler wobble, due to the differential rotation and the resulting centrifugal potential, then has a very small effect on the mean inertia moments and almost does not change the effective flattening. In eqn [47] for the Chandler wobble period, the term k_2/k is then very small compared to 1. The slow rotation of Venus also implies a much less efficient damping than on Earth (Yoder and Ward, 1979). A larger CW amplitude than for the Earth would then be possible even for a smaller excitation rate than that required to maintain the Earth's CW. Excitation could be due to seismic activity, mantle convection, or atmosphere/climate variations. By assuming quake-induced ΔC_{21} comparable to the Earth, Yoder and Ward (1979) estimated that a wobble amplitude $< 1^\circ$ requires a seismic activity rate on Venus $\leq 10^{-2}/Q(\sigma_{\mathrm{CW}})$ times the rate of seismic activity on the Earth. The small observed wobble suggests that Venus is significantly less active than the Earth (Yoder, 1995, 1997).

Short periodic LOD variations and orientation changes in space (nutation) have not yet been observed and are expected to be small. Because of the almost perpendicular position of Venus' rotation axis with respect to the orbital plane (see **Table 1**), Venus does not have seasons as on the Earth and Mars, for which the seasonal atmospheric cycle is the main driver of LOD variations. Periodic LOD variations of Venus have been calculated theoretically from GCMs of the atmosphere of Venus, which show that changes in the angular momentum of Venus are of the same order of magnitude as on Earth but have a main period equal to the solar day of 117 days as a result of variations in the received solar radiation (Karatekin et al., 2011; Lebonnois et al., 2010). Since the solid-body angular momentum of Venus is two orders of magnitude smaller than for the Earth because of Venus' slower rotation, larger rotation variations might be expected on Venus, but the induced surface displacements are only a few meters because of this slow rotation. The LOD variations are mainly driven by the wind, as for the Earth but in contrast with Mars (Karatekin et al., 2011).

Nutations on Venus are expected to be somewhat smaller than on Earth. The solar torque, which causes the nutations, is much smaller than for the Earth mainly because the polar flattening is a factor 250 smaller (Konopliv and Yoder, 1996). The differences with respect to the Earth due to Venus' small obliquity and shorter distance to the Sun are much less significant. The shorter distance to the Sun would tend to increase the nutations by a factor 2.6, and the small obliquity (see the $\sin 2\delta$ in eqns [65] and [66]) reduces the nutation amplitudes with respect to the Earth by a factor of 8. On the other hand, the slow rotation of Venus implies a weaker rotational stability, and orientation changes such as nutation are larger for a given torque (see also eqns [63] and [64]), which almost compensates for the smaller torque. The small eccentricity of the orbit of Venus suggests that nutations other than the semiannual nutation would be smaller.

For the Earth and Mars, nutation observations have and can be used to determine properties of the deep interior of these planets. For Venus, however, nutations observations have less geophysical potential because resonance effects with the FCN are expected to be negligible. The FCN period is about 10^5 years in inertial space (we assumed a similar CMB flattening as for the total planet and neglected the very small effect of deformation), far from the main nutation excitation periods, which are associated with Venus' orbital motion.

10.04.5 Mercury

10.04.5.1 Current Rotation State

Based on radar observations of the rotation of Mercury made at the Arecibo Observatory in Puerto Rico (Pettengill and Dyce, 1965), Colombo (1965) proposed the idea that Mercury has a stable rotation period exactly equal to 2/3 of the orbital period. This 3:2 spin–orbit resonance is now universally accepted. Before, it was thought that Mercury's spin was in a 1:1 resonance with its orbital motion, just like the Moon in its orbit around the Earth (see, e.g., Chapman, 1988). The rotation period of Mercury is 58.64601 days (Margot, 2009).

Mercury closely occupies the Cassini state 1, in which the rotation axis and the orbit normal remain coplanar with the normal to the Laplace plane as the rotation axis of Mercury precesses about the orbit normal and the orbit normal precesses about the normal to the Laplace plane (Colombo, 1966; Margot et al., 2007; Peale, 1969; Ward, 1975). The Laplace plane is defined as the plane about which Mercury's orbit precesses with constant inclination between the two planes. A Cassini state represents an extreme (equilibrium point) in the Hamiltonian of motion, made integrable by limiting the planetary perturbations (mainly due to Venus, the Earth, and Jupiter) to a single term corresponding to a regular motion of the ascending node, and has first been determined empirically for the Moon (Cassini, 1693). Observational evidence for this state has been obtained by Margot et al. (2007). From the observations of the determination of the epoch and time delay that maximize the cross correlation of the backscattered radar signal at two antennas separated by a long baseline (Holin, 1999), Margot et al. (2012) had shown that the spin pole position differs by 2.7 arcsec from the prediction Cassini 1 position, within the 5 arcsec uncertainty on the orientation. The obliquity of Mercury is 2.4 ± 0.08 arcmin (Margot et al., 2012).

10.04.5.2 Long-Term Spin Evolution

10.04.5.2.1 Spin–orbit resonance

Because of Mercury's closeness to the Sun, tidal friction can slow down the rotation of Mercury to rotation periods

commensurate with the orbital period on a timescale much smaller than the age of the solar system (see eqn [58]). Correia and Laskar (2004) estimated that the time needed to despin Mercury from an initial rotation period of 10 h would be about 300 Ma for a Q-value of 50. The final equilibrium rotation period for Mercury is different from the rotation synchronous with the orbital motion expected for tidal despinning of planets in circular orbits because of Mercury's substantial eccentricity. If the rotation were synchronous, the angular rotation rate would be smaller than the orbital velocity close to perihelion and larger near to aphelion, resulting in accelerating and decelerating tidal torques, respectively. Because of its distance dependence, the tidal torque is larger near perihelion than at aphelion, and the mean tidal torque would act to spin up Mercury. For the present eccentricity ($e=0.206$), the equilibrium rotation velocity is $\omega=1.25685n$, and the present mean rotation rate of Mercury ($3n/2$) corresponds to the tidal equilibrium rotation rate for eccentricity $e_{3/2}=0.284927$ (Correia and Laskar, 2004, who assume the averaged tidal torque to be linear in the tidal frequency). The tidal torque equilibrium is currently not satisfied because Mercury is locked in the 3:2 spin–orbit resonance.

We here assume zero obliquity (the observed value is (2.04 ± 0.08) arcmin, Margot et al., 2012) and principal axis rotation for the study of Mercury's rotation rate. Only the z-component of the Liouville equations must then be considered. For an entirely solid and rigid planet, eqn [25] reduces to

$$C\Omega\dot{m}_z = L_P \qquad [79]$$

where

$$L_P = \frac{3}{2}(B-A)\frac{GM_\odot}{r^3}\sin 2\xi \qquad [80]$$

is the solar torque on Mercury's permanent figure. We for the moment neglect the tidal torques. The distance r is the distance between the mass centers of the Sun and Mercury, and ξ is the angle between the direction of Mercury's axis of smallest moment of inertia A (also called the long axis) and the direction to the Sun:

$$\xi = f - \phi \qquad [81]$$

where f is the true anomaly and ϕ the rotation angle between the long axis of Mercury and the major axis of Mercury's orbit considered fixed (see Figure 3). We consider a rotation close to the 3:2 spin–orbit resonance and introduce the angle $\gamma=\phi-\frac{3}{1}M$, where M is the mean anomaly. At perihelion passage, γ is the angle between the direction of Mercury's long axis and the direction to the Sun. From the definitions of the angles γ and ξ, we have

$$\xi = f - \gamma - \frac{3}{2}M \qquad [82]$$

and the torque equation can be expressed as

$$C\ddot{\gamma} = -\frac{3}{2}n^2(B-A)\frac{a^3}{r^3}\sin(2\gamma+3M-2f) \qquad [83]$$

where the mean motion $n=\sqrt{G(M_\odot+M_{Merc})/a^3}\approx\sqrt{GM_\odot/a^3}$ is used. The time-dependent factor $\frac{1}{r^3}\sin(2\gamma+3M-2f)$ of the torque can be expanded in terms of the mean anomaly M as

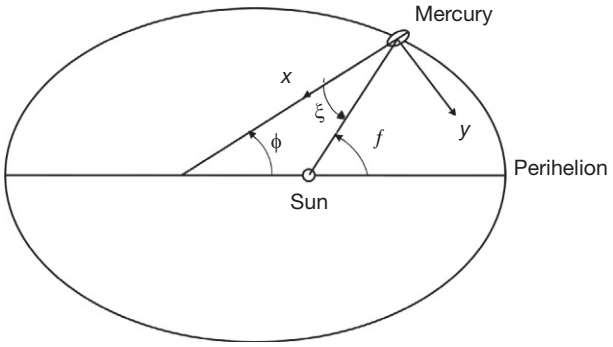


Figure 3 Geometry of Mercury's spin–orbit motion. Mercury's long axis makes an angle ξ with respect to the Sun and an angle ϕ with respect to an inertial reference axis, here taken to be temporarily coinciding with the semimajor axis of Mercury's orbital motion around the Sun.

$$\frac{1}{r^3}\sin(2\gamma+3M-2f) = \frac{1}{a^3}\sum_q G20_q(e)\sin[2\gamma+(1-q)M] \qquad [84]$$

(e.g., Murray and Dermott, 1999), where $G_{20q}(e)$ are eccentricity functions (Kaula, 2000). By developing the sine terms and retaining only terms up to e^4 in the eccentricity functions, we have

$$\begin{aligned}&\frac{d^2\gamma}{dM^2}+\frac{3}{4}\frac{B-A}{C}e\left(7-\frac{123}{8}e^2\right)\sin 2\gamma\\&=-\frac{3}{2}\frac{B-A}{C}\left[\left(1-11e^2+\frac{959}{48}e^4+\cdots\right)\sin M\right.\\&\left.-\frac{1}{2}e\left(1+\frac{421}{12}e^2+\cdots\right)\sin 2M+\cdots\right]+\cdots\end{aligned} \qquad [85]$$

(Peale, 2005). The angle γ is considered very small, and time-dependent forcing terms proportional to sin 2γ have been neglected compared to the other terms in the right-hand side.

For the study of the stability of the 3:2 resonance, eqn [85] is averaged over an orbital period. Since the torques are small, ξ does not change much over an orbital period, and $\dot{\xi}$ (or $\dot{\gamma}$) can be kept constant in the averaging procedure. The periodic forcing terms in the right-hand side of eqn [85] then vanish, and the equation reduces to the harmonic oscillator (or pendulum) equation:

$$\frac{d^2\gamma}{dM^2}+\frac{3}{4}\frac{B-A}{C}e\left(7-\frac{123}{8}e^2\right)\sin 2\gamma = 0 \qquad [86]$$

which has $\gamma=0$ as a stable solution. This solution corresponds to the long axis of Mercury pointing toward the Sun at perihelion.

The effect of tidal dissipation can be included by adding the averaged tidal torque $<L_T>$ divided by n^2C to the right-hand side of eqn [85]. If $<L_T>$ can be assumed to be constant (independent of $\gamma,\dot{\gamma},t$), the stability of the 3:2 resonance is maintained if the averaged tidal torque does not exceed the maximum possible restoring torque:

$$|<L_T>| < \frac{3}{4}(B-A)n^2e\left(7-\frac{123}{8}e^2\right) \qquad [87]$$

(Colombo and Shapiro, 1966; Goldreich and Peale, 1966b). By using eqn [56], this translates into a condition on the permanent deformation and the tidal dissipation factor Q:

$$\frac{B-A}{4MR^2} > \frac{k_2}{14Q}\left(\frac{R}{a}\right)^3 \frac{M_\odot}{M}\frac{1+\frac{123}{56}e^2}{e} \approx \frac{7\times 10^{-8}}{Q}, \qquad [88]$$

where we used $k_2 \approx 0.4$ (Rivoldini et al., 2009; Spohn et al., 2001; Van Hoolst and Jacobs, 2003; Van Hoolst et al., 2007). Given that the moment of inertia difference $B-A = (3.235\pm0.024)\times10^{-5}MR^2$ (Smith et al., 2012) and that Q is in the range from 10 to 500 (Goldreich and Soter, 1966), this stability condition is certainly satisfied. For a circular orbit, the 3:2 resonance would not be stable. Note that a large equatorial flattening and a large eccentricity contribute to the stability of the resonance. Several other resonances in the vicinity of the 3:2 resonance are also stable (Goldreich and Peale, 1966b). Why then did Mercury select the 3:2 resonance?

A necessary condition for capture is that the averaged tidal torque is not symmetrical about $\dot{\gamma}=0$; otherwise, a decelerating Mercury would continue to decrease its rotation rate (Goldreich and Peale, 1966b; Peale, 1976a). If the phase lag is proportional to the tidal frequency (Q inversely proportional to tidal frequency), this condition is satisfied. Even for constant Q dissipation models, the torque depends on the sign of the phase angle (or the sign of $\dot{\gamma}$, e.g., Kaula, 1964), and it can be concluded that the averaged tidal torque is expected to be asymmetrical about $\dot{\gamma}$ for all tidal models (Goldreich and Peale, 1966b). The precise capture probability is quite model-dependent. Assuming the dissipation factor to be inversely proportional to the frequency, Goldreich and Peale (1966b) found that Mercury has the largest probability to be captured in the 3:2 resonance, with a probability of 6.7%. The capture probability is rather small but can be increased if dissipative core–mantle coupling is included. Other resonances will then also have increased capture probabilities, and specific core viscosity values are needed to avoid capture in the 2:1 resonance (Peale and Boss, 1977). In these calculations, the present value of the eccentricity was used. However, Mercury's eccentricity shows large chaotic variations in time between zero and about 0.5 due to planetary influences (Laskar, 1994). As a result, over time, Mercury's eccentricity could have passed several times through the value $e_{3/2}$ at which tidal torques tend to drive Mercury's rotation velocity to the 3:2 resonant value, thereby considerably increasing the capture probability (Correia and Laskar, 2004). Moreover, Mercury could have escaped from an initial resonance capture when the eccentricity decreased to low values close to zero. The probability that the 3:2 spin–orbit resonance is the final state is 55.4% for a solid core (Correia and Laskar, 2004) and reduces to 22.6% for a liquid core (Correia and Laskar, 2009). However, large impacts may have disrupted any initial resonant capture and changed the final rotation state. The asymmetrical distribution of large impact basins is consistent with a previous synchronous rotation of Mercury, which is the most likely outcome for an initial retrograde rotation of Mercury. By including the effects of impacts, Correia and Laskar (2012) had shown that for both prograde and retrograde initial rotations, the 3:2 resonance is the most likely end state with a probability of around 50% and that the synchronous state is the second most probable.

10.04.5.2.2 Precession about the Cassini state and nutation

Tidal friction drives the spin of Mercury to Cassini state 1 from almost any initial condition on a timescale that is short compared to the age of the solar system (Peale, 1974). The primordial obliquity of Mercury was probably below 90°; otherwise, core–mantle friction could have driven Mercury to the Cassini state 3, similar to Cassini state 1 but with an obliquity 180° larger (Peale, 1976a). Provided that Mercury was born as a fast rotator, it must have passed a large chaotic zone in which the obliquity shows large and chaotic variations of tens of degrees within a few million years (Laskar and Robutel, 1993). The obliquity of the final evolved Cassini state 1 is given by

$$\frac{C}{MR^2} = \frac{\left[\frac{J_2}{(1-e^2)^{3/2}} + eC_{22}\left(7-\frac{123}{8}e^2\right)\right]}{\frac{\sin I}{\varepsilon_C}+\cos I}\frac{n}{\mu} \qquad [89]$$

(Peale, 1969), where ε_C is the obliquity of Cassini state 1, $I=8.6°$ the inclination of the orbital plane to the Laplace plane (Seidelmann, 2005), and $\mu = 2\pi/(328\,000$ years) the precession rate of Mercury's orbit about the normal to the Laplace plane (Yseboodt and Margot, 2006). Because the rotation axis is coplanar with the orbit normal and the normal to the Laplace plane in a Cassini state, the obliquity of the rotation axis is constant. For the recently determined degree-two gravitational coefficients from MESSENGER radio tracking ($J_2 = (5.03\pm0.02)\times10^{-5}$ and $C_{22} = (0.809\pm0.006)\times10^{-5}$, Smith et al., 2012) and the observed obliquity of (2.04 ± 0.08) arcmin (Margot et al., 2012), eqn [89] shows that $C/MR^2 = 0.346\pm0.014$, putting a strong constraint on Mercury's interior structure (Hauck et al., 2013; Rivoldini and Van Hoolst, 2013; Smith et al., 2012). Note that the large C_{22} and J_2 values compared to those expected for a hydrostatic Mercury with the current rotation rate and 3:2 spin–orbit resonance could be explained by a large remnant bulge caused by an initial faster rotation and subsequent polar wander due to the formation of the Caloris impact basin (Matsuy-Mama and Nimmo, 2009).

The precession period of both the orbit normal and the rotation axis around the normal to the Laplace plane is about 328 000 years (Yseboodt and Margot, 2006). Besides the slow precession of the orbit normal and the rotation axis around the normal to the Laplace plane, the rotation axis of Mercury also precesses about its position in the Cassini state if Mercury does not exactly occupy but is close to the Cassini state (Peale, 1974; Rambaux and Bois, 2004; Ward, 1975). If we neglect the much slower orbital precession, the latter precession due to the solar torque (also called free precession) can be compared to the precession of the Earth about the normal to the ecliptic. However, due to Mercury's 3:2 spin–orbit resonance, the classical analytic expression for the precession rate (see, e.g., Smart, 1953) must be extended with a term in the equatorial flattening (or C_{22}). By averaging the Liouville equations without deformation effects over the orbital period, the precession period can be derived, correct up to order e^3, as

$$P_{\text{Prec}} = \frac{3\pi C^m}{nMR^2}\left\{\left[\frac{3J_2}{2(1-e^2)^{3/2}} + \frac{3}{2}eC_{22}\left(7-\frac{123}{8}e^2\right)\right]^2 - C_{22}^2\frac{159}{16}e^6\right\}^{-1/2} \quad [90]$$

(Peale, 2005). The C_{22} terms decrease the traditional precession period by nearly 30% and cause the precession motion to be slightly elliptical (Peale, 2005). For a solid core, $C = 0.346MR^2$, and the MESSENGER gravitational coefficients (Smith et al., 2012), the precession period is 1297 years. However, at least the outer part of Mercury's core is liquid as follows from Mercury's libration observations (Margot et al., 2007; see next section) and $C^{\text{m}}/(MR^2)$ is estimated to be 0.148 ± 0.006 (next section). The precession period is then estimated to be 556 years, shorter than for an entirely solid Mercury, as expected for short periodic mantle precession.

Tidal friction and dissipative core–mantle coupling will damp free precession, that is, will reduce the obliquity variations, and will drive the rotation axis to the Cassini state if the orbital inclination to the invariable plane is fixed and Mercury precesses about the Cassini state at constant rate (Peale, 1974, 2005; Ward, 1975). Because of planetary perturbations, the precession rate and inclination show large variations, which suggests that Mercury cannot remain exactly in the Cassini state. However, the precession period is much smaller than the shortest timescales of about 10^4 to 10^6 years for the variations in the planetary orbital parameters. Therefore, the rotation axis accurately tracks the changes in the orientation of the orbit normal. If Mercury has once been driven close to the Cassini state by dissipative effects, it is expected to remain very close to the instantaneous location of the Cassini state (Goldreich and Toomre, 1969; Peale, 1974). With a typical core viscosity value for the Earth's core determined from ab initio calculations (de Wijs et al., 1998) and $Q \approx 100$, Peale (2005) obtained free precession damping timescales of about 10^5 years. This is much shorter than the age of the solar system, suggesting that free precession is either completely damped or recently excited (Peale, 1974; Ward, 1975). On the other hand, the damping timescale is somewhat longer than the 10^4-year timescale of the Laplace plane reorientation, and the changing Cassini state may never be reached exactly, but nevertheless, the rotation axis remains close to the Cassini state, even if the Cassini obliquity changes due to the precession of the orbit (Yseboodt and Margot, 2006). Peale (2006) followed the spin position and the Cassini state position during short timescale orbital variations over 20 000 years as well as during long timescale variations over the past 3 My and showed that the spin axis remains within one arcsec of the Cassini state if it initially occupied the Cassini state. On the other hand, Bills and Comstock (2005) obtained obliquity variations forced by the orbital motion on timescales of My with typical obliquity values of 2–4 arcmin, values that are larger than the observed obliquity. The obliquity history looks very much like a scaled version of the inclination history, with the obliquity variations being roughly 200 times smaller than the corresponding inclination variations and typical periods of about 100 ky and 1 My (Bills and Comstock, 2005).

Another source of possible deviations from the Cassini state is short periodic variations in obliquity and longitude caused by the solar torque. These variations, similar to the nutations for the Earth and Mars, have main frequencies equal to the harmonics of the orbital frequency (Carpentier and Roosbeek, 2003). The amplitudes are on the order of 0.1 arcsec or smaller (Carpentier and Roosbeek, 2003; Rambaux and Bois, 2004). They have been calculated for a rigid Mercury, but are not expected to be largely amplified when the interior properties of Mercury are taken into account. For the Earth and Mars, the FCN frequency is close to the main nutation frequencies, and some nutations are resonantly amplified. For Mercury, the FCN period is at least several hundred years, as can be estimated from eqn [48], and is far from the period of the nutations.

10.04.5.3 Librations

10.04.5.3.1 Forced libration

Because of the gravitational torque on the permanent figure of Mercury in its elliptical orbit, the rotation of Mercury is not constant and varies about a mean state. For the analysis of these librations, we include the liquid core, but still neglect deformation (no incremental inertia terms). Liouville eqns [25] and [37] can then be written as

$$C\Omega\dot{m}_z + C^{\text{f}}\Omega m_z^{\text{f}} = \frac{3}{2}(B-A)\frac{GM_\odot}{r^3}\sin 2\xi \quad [91]$$

$$C^{\text{f}}\Omega\left(\dot{m}_z + \dot{m}_z^{\text{f}}\right) = 0 \quad [92]$$

where the last equation indicates that the core does not follow the rotation variations, or librations, of the mantle. By substituting eqn [92] into eqn [91], and taking into account that $C = C^{\text{m}} + C^{\text{f}}$, the Liouville equation for the whole planet can be transformed to a Liouville equation for the mantle:

$$C^m\ddot{\gamma} = \frac{3}{2}(B-A)\frac{GM_\odot}{r^3}\sin 2\xi \quad [93]$$

where the libration γ is now defined in terms of the direction of the long axis of the mantle. With inclusion of a liquid core, the only change in eqns [83] and [85] is that C must be replaced by C^{m}. Equation [85] with C replaced by C^{m} is an equation for a forced harmonic oscillator describing forced libration in longitude. The angle γ is usually called the libration angle. The forcing is at 1 orbital period, 1/2 orbital period, and higher subharmonics of the orbital period, which have been neglected here. Due to the small ratio $(B-A)/C^{\text{m}}$, the amplitude γ_1 of the 88-day libration is equal to the amplitude of the sin M term in eqn [85]:

$$\gamma_1 = \frac{3}{2}\frac{B-A}{C^{\text{m}}}\left(1 - 11e^2 + \frac{959}{48}e^4 + \cdots\right) \quad [94]$$

(Peale, 1972), and the amplitude of the sin $2M$ term is about 11% of that (Jehn et al., 2004; Van Hoolst et al., 2012). Because of its dependence on the mantle moment of inertia, the forced libration amplitude offers a unique opportunity to study the deep interior of Mercury. Mantle must be understood here as the solid part of the planet above the liquid core that partakes in the libration and thus also includes the crust.

Margot et al. (2012) determined the amplitude of the 88-day libration to be (38.5 ± 1.6) arcsec, which corresponds

to a maximum displacement at the equator with respect to the equilibrium rotation of (455 ± 19) m. For the MESSENGER $B-A$ value (Smith et al., 2012), eqn [94] shows that $C^m/(MR^2) = 0.148 \pm 0.006$ (Margot et al., 2012; Rivoldini and Van Hoolst, 2013), less than half the value for the total planetary moment of inertia $C/(MR^2) = 0.346 \pm 0.014$ determined from the obliquity value. Therefore, Mercury cannot be entirely solid, and a global liquid layer must exist relatively much closer to the surface than, for example, the fluid outer core of the Earth that is separated from the surface. The comparison of the rotational constraints on C and C^m with theoretical moment of inertia values for models of the interior structure of Mercury shows that the liquid core has a radius of about 2000 km (Hauck et al., 2013; Rivoldini and Van Hoolst, 2013; Smith et al., 2012).

For a correct interpretation of libration observations, all possible effects on the rotation angle variations of Mercury have to be quantified, such as the effect of neglected deformation and core–mantle couplings in eqn [94]. These effects have been shown to be important for changes in the rotation angle of the Earth. For example, on a decadal timescale, the changes in the rotation angle of the Earth are attributed to the exchange of angular momentum between the core and the solid mantle (Jault et al., 1988), but there is no general agreement on which physical mechanism dominates.

Electromagnetic coupling between the core and the mantle arises from the flow of electric current in the weakly conducting mantle. The mantle conductivity is assumed to be concentrated in a thin layer at the CMB, possibly resulting from chemical reactions between the mantle silicates and the conducting liquid iron assemblage (Buffett et al., 2000). Holme (1998) showed that electromagnetic coupling is capable of explaining the observed changes in the rotation rate of the Earth. Peale et al. (2002) applied Buffett's (1992) formalism for electromagnetic coupling to Mercury and estimated the torque related to the poloidal magnetic field from Mercury's observed magnetic dipole moment of $6 \times 10^{12}\,\mathrm{Tm}^3$ as estimated from Mariner 10 data (Russell et al., 1988; note that the new MESSENGER estimate of Anderson et al., 2012, does not affect the conclusions). The time constant for the decay of the relative angular velocity between the core and the mantle due to the electromagnetic torque is estimated to be on the order of 10^6 years. As this is much larger than the 88-day period of the forced libration, the electromagnetic torque does not affect the forced libration (Peale et al., 2002). The torque could, however, be larger, and the decay time smaller, since only the observed dipole moment is used in the calculation and the torque involving the toroidal magnetic field has been neglected. As the dynamo action in Mercury is not well understood and different mechanisms have been proposed (e.g., Christensen, 2006; Manglik et al., 2010; Takahashi and Matsushima, 2006; Vilim et al., 2010), it cannot be excluded that the dipole field of Mercury is even very small compared to the total magnetic field in the core. For example, Stanley et al. (2005) showed that thin shell dynamos, which could be applicable to Mercury, can produce magnetic fields of different geometry than the Earth's magnetic field and with Mercury-like dipolar field intensities. The toroidal field strength can be 100 times larger than the dipolar strength, and the poloidal field strength can be 10 times larger than the dipolar strength. Projecting the dipole field observed by Mariner 10 and MESSENGER to the CMB then does not provide a good estimate for the CMB field.

Pressure coupling between the liquid core and the mantle requires a nonspherical interface. Due to the 3:2 spin–orbit resonance, static tides will induce a small difference in the long and short equatorial axes. These differences are, however, very small and of the order of a meter (Matsuyama and Nimmo, 2009). Mantle convection (e.g., Schubert et al., 1988), if it still is operational (Grott et al., 2011; Michel et al., 2012; Reese and Peterson, 2002), induces further topography on the CMB. In the thin Mercurian mantle, convection cells with relatively small horizontal length scales comparable to the mantle depth become unstable for the lowest Rayleigh numbers (Schubert et al., 2001). The resulting deviations from the hydrostatic equilibrium form are not expected to contribute largely to B^f-A^f, although some models suggest that convection might be organized in convection cells of large horizontal scale (Conzelmann and Spohn, 1999). As the shape of the CMB is essentially unknown, Peale et al. (2002) rescaled the topographic torque on the Earth's mantle needed to explain the Earth's LOD variations to Mercury in order to estimate the effect of topographic pressure coupling on Mercury's librations and showed that the change in the forced libration is essentially undetectable (Peale et al., 2002). Inertial coupling on the nonspherical CMB also has a negligible effect (Rambaux et al., 2007), and viscous coupling at the CMB can be neglected too since the relaxation timescale for relative motion between the mantle and the liquid core is several orders of magnitude longer than the 88-day period of libration for any reasonable value for the core viscosity.

An inner core can also interact directly with the mantle through gravitational coupling if the long axes of inner core and mantle are not aligned and influence the mantle libration. The global magnetic field observed by Mariner 10 and MESSENGER, if caused by dynamo action, not only is further evidence for a liquid outer core but also suggests that Mercury contains a growing solid inner core (Schubert et al., 1988; Spohn, 1991; Stevenson et al., 1983). The inner core is possibly even relatively large (Grott et al., 2011; Hauck et al., 2004; Stevenson et al., 1983) although its size is limited by the global radial contraction of 2.4–3.6 km estimated from Mercury's tectonics (Di Achille et al., 2012). The effect of an inner core on Mercury's 88-day libration amplitude has been shown to be below the observational precision for small inner cores but could be observed with the current observational uncertainty if Mercury's inner core is larger than at least 1000 km (Van Hoolst et al., 2012). Gravitational inner core–mantle coupling slightly reduces the mantle libration by up to 20 m, and therefore, the mantle moment of inertia derived from the libration is also somewhat smaller than if derived from eqn [94] (Van Hoolst et al., 2012).

Tides also affect the librations by changing the gravitational torque exerted by the Sun on Mercury and by changing the polar moment of inertia. Their effect on the amplitude of the 88-day forced libration is, however, at most 0.5% (Van Hoolst et al., 2012), corresponding to a difference of below 2 m, which is below the (future) observational precision expected to be about 10 m or somewhat smaller (Pfyffer et al., 2011). The main cause of the difference is the change in the torque (Van Hoolst et al., 2012).

10.04.5.3.2 Free libration

Libration (eqn [85]) averaged over an orbital period governs free libration. The periodic torques with periods equal to the mean motion and its harmonics, the terms on the right-hand side of eqn [85], are therefore not considered for free libration. The total torque of the Sun is, however, not neglected; otherwise, there would not be any libration anymore. The part of the torque that causes free libration is proportional to the libration angle γ (see the second term in the left-hand side of eqn [85]). When the long axis does not point toward the Sun at perihelion ($\gamma \neq 0$), the averaged gravitational torque on the permanent figure of Mercury tends to restore the alignment, and the long axis will librate around the direction to the Sun at perihelion. The oscillation period is given by

$$P_{\text{free}} = \frac{2\pi}{n}\left[\frac{1}{3}\frac{C^{\text{m}}}{B-A}\frac{1}{e\left(\frac{7}{2}-\frac{123}{16}e^2\right)}\right]^{1/2}. \qquad [95]$$

Free libration is also due to the gravitational interaction between the Sun and Mercury, as is forced libration, but the distinctive feature of free libration as opposed to forced libration is that the amplitude and phase of free libration cannot be determined without knowing its excitation and dissipation, as is the case for a free harmonic oscillator. For $B-A=(3.235\pm0.024)\times10^{-5}MR^2$ (Smith et al., 2012) and $C^{\text{m}}/(MR^2)=0.149\pm0.006$ (Margot et al., 2012), the free libration period is 11.4 years. Although the long period makes this effect more difficult to detect in a few orbital periods of Mercury, the expected lifetime of the space missions MESSENGER and BepiColombo, it also causes the free libration to be much more difficult to detect in spin rate measurements, as currently done with Earth-based radar, than in measurements of the rotational phase with camera, altimeter, or gravitational sensors (Yseboodt et al., 2010).

The amplitude of the free libration depends on excitation and damping processes. Free libration will eventually be damped to zero in the absence of excitation. A main source of damping is the tides caused by the Sun. Mercury reacts viscoelastically to forces that tend to deform it, and inevitably, some energy is lost. A further potential source of damping is due to dissipative core–mantle coupling, due to either core viscosity or ohmic dissipation in the core. The timescale for damping of the free libration by viscous core–mantle coupling and tides is on the order of 10^5 years (Peale, 2005). For the same dissipation parameters as used for the damping of the free precession, the viscous core–mantle coupling dominates the tidal damping by one order of magnitude (Peale, 2005). The damping timescales are short compared to the age of the solar system. Without recent excitation, the free libration is therefore expected to be completely damped. Impact excitation is very unlikely since the average time span between impacts of sufficient size is about 10^9 years with current cometary fluxes. A recent impact would have left an impact crater of at least 20 km diameter for the excitation of an observable amplitude of 0.1 arcmin (Peale, 2005).

Librations at periods longer than the orbital period are also induced by planetary perturbations as they change Mercury's orbit (Peale et al., 2007). For example, long-period changes on the order of several years in the semimajor axis of Mercury's orbit caused by the other planets imply a change in the mean motion and cause the long axis of Mercury (see libration angle γ) to be misaligned with respect to the direction to the Sun at pericenter. The solar gravitational torque on Mercury then leads to forced librations at these periods. Several of those long-period librations can have amplitudes above 1 arcsec (Peale et al., 2009; Yseboodt et al., 2010, 2013). The largest long-period libration most likely is caused by Jupiter and has a period of 11.86 years. This forcing period is close to the free libration period of 11.4 years. Although it is, within the current error estimates, not equal to the free period, the near resonance can amplify the libration response of Mercury.

When Mercury has a solid inner core inside an outer liquid core, the free libration period can be considerably lengthened (Dumberry, 2011) to a period of about 17.4 years for solid inner cores almost as large as the total core (Van Hoolst et al., 2012). For small inner cores (radius of less than about 1000 km), this can bring the free mode closer to resonance with the Jupiter-induced long-period libration causing a larger resonant amplification. For larger inner core, the free libration period is far from 11.86 years. An inner core changes the free libration period because the inner core librates in phase with the mantle. In addition to changing the free libration, the existence of a solid inner core also introduces a second free mode in which the mantle and inner core librate out of phase (Dumberry, 2011). This second period is shorter than the period of the first free mode, of the order of several years, and might resonantly enhance other long-period librations (Van Hoolst et al., 2012; Yseboodt et al., 2013). Future observations of one of the free modes or of the amplification of a long-period forced libration can yield constraints on the interior structure of Mercury, including its inner core size.

Besides excitation of long-period libration due to planetary perturbations, decadal libration might also be caused by time-dependent zonal flows generated by turbulent convection in Mercury's fluid outer core (Koning and Dumberry, 2013). Those flows periodically speed up and slow down the inner core through electromagnetic interaction so that also the mantle will start librating as a result of gravitational interaction with the inner core. The libration is largest at forcing periods close to the free libration period at decadal period, and libration amplitudes of the order of 1 arcsec can be reached for zonal flows of the order of 1 km year^{-1} (Koning and Dumberry, 2013). Future observations of decadal libration could be used to constrain the dynamics in the core of Mercury.

10.04.5.4 Wobble

Polar motion of Mercury has never been detected. Due to the very tenuous atmosphere of Mercury, seasonal atmospheric excitation of polar motion as on Earth can be neglected. As for Venus, the CW component of polar motion has a very long period because of Mercury's slow rotation and small flattenings, which are two orders of magnitude larger than the polar rotational flattening. No strong excitation is known at this timescale. Seismic activity is a potential source of CW excitation but is expected to be orders of magnitude smaller than on Earth for the lack of plate tectonics. On the other hand, the dissipation of the CW is less efficient than on Earth due to the slow rotation of Mercury but, for weak excitation, will eventually damp the wobble to very small values.

10.04.6 Summary

The rotation of the four terrestrial planets of our solar system has changed substantially since the planets' formation. The rotation history cannot be uniquely described, but recent advances have led to a deeper understanding of their current rotation state. All four planets could have experienced large chaotic variations in obliquity due to resonances between the precession rate and the orbit precession rate. Their obliquities can therefore not be considered as primordial and almost do not constrain the primordial obliquities. The rotation rates have also been changed. Because Mars is further from the Sun than the other terrestrial planets and has no large satellites, tidal braking has only had a small influence on its spin rate, which can be considered to be primordial. Mercury and Venus are in a stable final rotation state, as a result of dissipative processes. Therefore, their current spin gives little insight into their primordial rotation. Although the occupied final spin states have been shown to be the most probable outcomes of the rotational evolution, the physical dissipation mechanisms are not well constrained and need further study.

Short periodic rotation variations of the Earth have been studied since antiquity and are nowadays routinely tracked with submilliarcsecond accuracy (see **Chapters 3.10** and **3.09**). Theoretical techniques developed for the Earth are now being applied to the other terrestrial planets, but data have so far only been obtained for the largest rotation variation components with limited precision. Seasonal LOD variations on Mars have recently been determined with a precision of about 5% for the largest signal and can be used to constrain GCMs of the Martian atmosphere and the seasonal sublimation/condensation cycle of the polar caps. By using Earth-based radar observations, both the libration and the obliquity of Mercury have recently been measured with an uncertainty of also about 5%. These data prove that the core of Mercury is at least partially liquid and offer unprecedented constraints on the deep interior structure of Mercury.

The next decade promises to be an exciting one for the study of periodic rotations variations of Mars and Mercury. Continued radio-tracking observations of Mars orbiters such as Odyssey and MRO will improve LOD variations and make it possible to resolve the Mars CW period. This will aid in the construction of accurate GCMs and will constrain further the interior of Mars. In addition, in 2016 the NASA Discovery Program mission InSight (Interior Exploration using Seismic Investigations, Geodesy and Heat Transport) will place a geophysical lander on Mars to study its deep interior. The RISE experiment of InSight will use the X-band communications transponder to obtain two-way Doppler and ranging measurements from the radio links resulting in an order of magnitude better accuracies on the rotation, on the order of a few mas. Besides improving LOD variations and polar motion, it will detect the nonrigid contributions to nutations and so offer direct information on the core of Mars. Earth-based radar observations of Mercury and measurements by the MESSENGER and BepiColombo missions to Mercury will improve the libration and obliquity determination to an accuracy of about an arcsec or better. The total planetary moment of inertia and the ratio of mantle to whole planet moment of inertia can then accurately be determined, with expected precisions of about 1%, yielding important information on Mercury's deep interior. Moreover, Venus Express data are expected to continue better constraining the rotation variations induced by the atmosphere.

Acknowledgments

The author thanks M. Dumberry and W. Folkner for their careful reviews.

References

Agnor CB, Canup RM, and Levison HF (1999) On the character and consequence of large impacts in the late stage of terrestrial accretion. *Icarus* 142: 219–237.

Aharonson O, Zuber MT, Smith DE, Neumann GA, Feldman WC, and Prettyman TH (2004) Depth, distribution, and density of CO_2 deposition on Mars. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2003JE002223, E05004.

Anderson DL (1989) *Theory of the Earth*. Boston, USA: Blackwell Scientific Publications.

Anderson BJ, JohnsonC L, Korth H, et al. (2012) The global magnetic field of mercury from MESSENGER orbital observations. *Science* 333: 1859–1862.

Andoyer H (1923) *Cours de Mecanique Celeste*. Paris: Gauthier-Villars.

Barnes RTH, Hide R, White HH, and Wilson CA (1983) Atmospheric angular momentum fluctuation, length-of-day changes and polar motion. *Proceedings of Royal Society of London Series A* 387: 31–73.

Barriot JP, Dehant V, Folkner W, et al. (2001) NEIGE: Net-lander ionosphere and geodesy experiment. *Advances in Space Research* 28(8): 1237–1249.

Bills BG (2005a) Free and forced obliquities of the Galilean satellites of Jupiter. *Icarus* 175: 233247.

Bills BG (2005b) Variations in the rotation rate of Venus due to orbital eccentricity modulation of solar tidal torques. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2003JE002190, E11007.

Bills BG and Comstock RL (2005) Forced obliquity variations of Mercury. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2003JE002116, E04006.

Bills BG, Kieffer WS, and Jones RLJ (1987) Venus gravity: A harmonic analysis. *Journal of Geophysical Research* 92: 232–250.

Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2004JE002376, E07004.

Bizouard C (1995) Modélisation astrométrique et géophysique de la rotation de la Terre. PhD thesis, Observatoire de Paris, France.

Borderies N (1980) Theory of Mars rotation in Euler angles. *Astronomy and Astrophysics* 82: 129–141.

Bouquillon S and Souchay J (1999) Precise modeling of the precession-nutation of Mars. *Astronomy and Astrophysics* 345: 282–297.

Bretagnon P and Francou G (1988) VSOP87: Planetary theories in rectangular and spherical variables. *Astronomy and Astrophysics* 202: 309–315.

Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics on Mars: Evidence from magnetic field history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072. http://dx.doi.org/10.1029/2002JE001999.

Buffett BA (1992) Constraints on magnetic energy and mantle conductivity from the forced nutations of the Earth. *Journal of Geophysical Research* 97(B13): 19581–19597.

Buffett BA and Christensen UR (2007) Magnetic and viscous coupling at the core-mantle boundary: Inferences from observations of the Earth's nutations. *Geophysical Journal International* 171: 145–152.

Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of the Earth's core. *Science* 290: 1338–1342.

Buffett BA, Mathews PM, and Herring TA (2002) Modeling of nutation and precession: Effects of electromagnetic coupling. *Journal of Geophysical Research*. http://dx.doi.org/10.1029/2000JB000056.

Busse FH (1968) Steady fluid flow in a precessing spheroidal shell. *Journal of Fluid Mechanics* 33: 739–751.

Carpenter RL (1964) Study of Venus by CW radar. *Astronomical Journal* 69: 2–11.

Carpenter RL (1966) Study of Venus by cw radar-1964 results. *Astronomical Journal* 71: 142–152.

Carpenter RL (1970) A radar determination of the rotation of Venus. *Astronomical Journal* 75: 61–66.

Carpentier G and Roosbeek F (2003) Analytical development of rigid Mercury nutation series. *Celestial Mechanics and Dynamical Astronomy* 86: 223–236.

Cassini GD (1693) Traité de l'origine et du Progrès de l'Astronomie. Paris: Académie Française.

Cazenave A and Balmino G (1981) Meteorological effects on the seasonal variations of the rotation of Mars. *Geophysical Research Letters* 8: 245–248.

Chambers JE (2001) Making more terrestrial planets. *Icarus* 152: 205–224.

Chao BF and Rubincam DP (1990) Variations of Mars gravitational field and rotation due to seasonal CO_2 exchange. *Journal of Geophysical Research* 95(B9): 14755–14760.

Chapman CR (1988) Mercury: Introduction to an end-member planet. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 1–23. Tucson, AZ, USA: Univ. Arizona Press.

Chapman S and Lindzen R (1970) *Atmospheric Tides*. Reidel, Dordrecht: Thermal and Gravitational.

Chapront-Touzé M (1990) Orbits of the Martian satellites from ESAPHO and ESADE theories. *Astronomy and Astrophysics* 240: 159–172.

Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.

Colombo G (1965) Rotational period of the planet Mercury. *Nature* 208: 575.

Colombo G (1966) Cassini's second and third laws. *Astronomical Journal* 71: 891–896.

Colombo G and Shapiro II (1966) The rotation of the planet Mercury. *The Astrophysical Journal* 145: 296–307.

Conzelmann V and Spohn T (1999) New thermal evolution models suggesting a hot partially molten Mercurian core. *Bulletin of the American Astronomical Society* 31: 1102.

Correia ACM and Laskar J (2001) The four final rotation states of Venus. *Nature* 411: 767–770.

Correia ACM and Laskar J (2003a) Long-term evolution of the spin of Venus II. Numerical simulations. *Icarus* 163: 24–45.

Correia ACM and Laskar J (2003b) Different tidal torques on a planet with a dense atmosphere and consequences to the spin dynamics. *Journal of Geophysical Research* 108(E11): 5123. http://dx.doi.org/10.1029/2003JE002059.

Correia ACM and Laskar J (2004) Mercury's capture into the 3/2 spin–orbit resonance as a result of its chaotic dynamics. *Nature* 429: 848–850.

Correia ACM and Laskar J (2009) Mercury's capture into the 3/2 spin–orbit resonance including the effect of core–mantle friction. *Icarus* 201: 1–11.

Correia ACM and Laskar J (2012) Impact cratering on Mercury: Consequences for the spin evolution. *The Astrophysical Journal Letters* 751: L43.

Correia ACM, Laskar J, and Néron de Surgy O (2003) Long-term evolution of the spin of Venus I. Theory. *Icarus* 163: 1–23.

Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ, USA: Princeton University Press.

Darwin GH (1908) *Tidal Friction and Cosmogony*, vol. 2. New York: Cambridge University Press, Scientific papers.

Davies ME, Colvin TR, Rogers PG, et al. (1992) The rotation period, direction of the North Pole, and geodetic control network of Venus. *Journal of Geophysical Research* 97(E8): 13141–13151.

de Wijs GA, Kresse G, Vočadlo L, Dobson D, Alfè Gillan MJ, and Price GD (1998) The viscosity of liquid iron at the physical conditions of the Earth's core. *Nature* 392: 805–807.

Defraigne P, de Viron O, Dehant V, Van Hoolst T, and Hourdin F (2000) Mars rotation variations induced by atmosphere and ice caps. *Journal of Geophysical Research* 105(E10): 24563–24570.

Defraigne P, Dehant V, and Paquet P (1995) Link between the retrograde-prograde nutations and nutations in obliquity and longitude. *Celestial Mechanics and Dynamical Astronomy* 62: 363–376.

Defraigne P, Dehant V, and Van Hoolst T (2001) Steady state convection in Mars' mantle. *Planetary and Space Science* 49: 501–509.

Defraigne P, Rivolidini A, Van Hoolst T, and Dehant V (2003) Mars nutation resonance due to free inner core nutation. *Journal of Geophysical Research* 108(E12): 5128. http://dx.doi.org/10.1029/2003JE002145.

Defraigne P and Smits I (1999) Length-of-day variations due to zonal tides for an inelastic earth in non-hydrostatic equilibrium. *Geophysical Journal International* 139: 563–572.

Dehant V, de Viron O, and Greff-Lefftz M (2005) Atmospheric and oceanic excitation of the rotation of a three-layer Earth. *Astronomy and Astrophysics* 446: 345–355. http://dx.doi.org/10.1051/0004-6361:20053825.

Dehant V, de Viron O, Karatekin Ö, and Van Hoolst T (2006) Excitation of Mars polar motion. *Astronomy and Astrophysics* 446: 345–355. http://dx.doi.org/10.1051/0004-6361:20053825.

Dehant V, Defraigne P, and Van Hoolst T (2000) Computation of Mars' transfer function for nutations, tides and surface loading. *Physics of the Earth and Planetary Interiors* 117: 385–395.

Dehant V, Folkner W, Renotte E, et al. (2009) Lander radio science for obtaining the rotation and orientation of Mars. *Planetary and Space Science* 57: 1050–1067.

Dehant V, Hinderer J, Legros H, and Lefftz M (1993) Analytical approach to the computation of the Earth, the outer core and the inner core rotational motions. *Physics of the Earth and Planetary Interiors* 76: 259–282.

Dehant V, Le Maistre S, Rivoldini A, et al. (2011) Revealing Mars' deep interior: Future geodesy missions using radio links between landers, orbiters, and the Earth. *Planetary and Space Science* 59: 1069–1081.

Dehant V, Van Hoolst T, de Viron O, Greff-Lefftz M, Legros H, and Defraigne P (2003) Can a solid inner core of Mars be detected from observations of polar motion and nutation of Mars? *Journal of Geophysical Research* 108(E12): 5127. http://dx.doi.org/10.1029/2003JE002140.

Dehant V, Van Hoolst T, and Defraigne P (2000) Comparison between the nutations of the planet Mars and the nutations of the Earth. *Surveys in Geophysics* 21: 89–110.

Deleplace B and Cardin P (2006) Viscomagnetic torque at the core mantle boundary. *Geophysical Journal International* 167(2): 557–566. http://dx.doi.org/10.1111/j.1365-246X.2006.03180.x.

Di Achille G, Popa C, Massironi M, et al. (2012) Mercury's radius change estimates revisited using MESSENGER data. *Icarus* 221: 456–460.

Dobrovolskis AR (1980) Atmospheric tides and the rotation of Venus. II. Spin evolution. *Icarus* 41: 18–35.

Dobrovolskis AR and Ingersoll AP (1980) Atmospheric tides and the rotation of Venus. I. Tidal theory and the balance of torques. *Icarus* 41: 1–17.

Dones L and Tremaine S (1993) On the origin of planetary spins. *Icarus* 103: 67–92.

Dreibus G and Wänke H (1985) Mars, a volatile-rich planet. *Meteoritics* 20: 367–381.

Dumberry M (2011) The free librations of Mercury and the size of its inner core. *Geophysical Research Letters* 38. http://dx.doi.org/10.1029/2011GL048277, L16202.

Dumberry M and Koot L (2012) A global model of electromagnetic coupling for nutations. *Geophysical Journal International* 191(2): 530–544. http://dx.doi.org/10.1111/j.1365-246X.2012.05625.x.

Dyce RB, Pettengill GH, and Shapiro II (1967) Radar determination of the rotations of Venus and Mercury. *Astronomical Journal* 72: 351.

Folkner WM, Kahn RD, Preston RA, et al. (1997) Mars dynamics from Earth-based tracking of the Mars Pathfinder lander. *Journal of Geophysical Research* 102(E2): 4057–4064.

Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from Radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.

Forget F, Hourdin F, Fournier R, et al. (1999) Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research* 104(E10): 24155–24176.

Getino J and Ferrándiz JM (2000) Effects of dissipation and a liquid core on forced nutations in Hamiltonian theory. *Geophysical Journal International* 142(3): 703–715.

Gold T (1955) Instability of the Earth's axis of rotation. *Nature* 175: 526529.

Gold T and Soter S (1969) Atmospheric tides and the resonant rotation of Venus. *Icarus* 11: 356–366.

Goldreich P and Peale SJ (1966a) Resonant rotation for Venus? *Nature* 209: 1117–1118.

Goldreich P and Peale S (1966b) Spin–orbit coupling in the Solar System. *Astronomical Journal* 71: 425–438.

Goldreich P and Peale SJ (1970) The obliquity of Venus. *Astronomical Journal* 75: 273–284.

Goldreich P and Soter S (1966) Q in the Solar System. *Icarus* 5: 375–389.

Goldreich P and Toomre A (1969) Some remarks on polar wandering. *Journal of Geophysical Research* 74: 2555.

Goldstein H (1950) *Classical Mechanics*. Reading, MA, USA: Addison-Wesley.

Goldstein RM (1964) Venus characteristics by Earth-based radar. *Astronomical Journal* 69: 12–18.

Golombek MP (2002) A revision of Mars seismicity from surface faulting. *Lunar and Planet Science Journal* XXXIII.

Golombek MP, Banerdt WB, Tanaka KL, and Tralli DM (1992) A prediction of Mars seismicity from surface faulting. *Science* 258: 979–981.

Gross R (2000) The excitation of the Chandler wobble. *Geophysical Research Letters* 27: 2329–2332.

Gross R, Fukumori I, Menemenlis D, and Gegout P (2004) Atmospheric and oceanic excitation of length-of-day variations during 1980–2000. *Journal of Geophysical Research B* 109(B1), B01406.

Grott M, Breuer D, and Laneuville M (2011) Thermo-chemical evolution and global contraction of mercury. *Earth and Planetary Science Letters* 307: 135–146.
Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of the Earth and Planetary Interiors* 142: 1–22.
Gwinn CR, Herring TA, and Shapiro II (1986) Geodesy by radio interferometry: Studies of the forced nutations of the earth 2. Interpretation. *Journal of Geophysical Research* 91(B5): 4755–4766.
Harder H and Christensen UR (1996) A one-plume model of Martian mantle convection. *Nature* 380: 507–509.
Hauck SA II, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222: 713–728.
Hauck SA, Margot JL, Sean C, et al. (2013) The curious case of Mercury's internal structure. *Journal of Geophysical Research: Planets* 118: 1–17.
Heiskanen WA and Moritz H (1967) *Physical Geodesy*. San Francisco: W. H. Freeman and Company.
Herschel W (1774) On the remarkable appearances of the polar regions of the planet Mars, the inclination of its axis, the position of its poles, and its spheroidal figure, with a few hints relating to its real diameter and atmosphere. *Philosophical Transactions* 24: 233–273.
Hilton JL (1991) The motion of Mars' pole. I. Rigid body precession and nutation. *Astronomical Journal* 102: 1510–1527.
Hilton JL (1992) The motion of Mars' pole. II. The effect of an elastic mantle and a liquid core. *Astronomical Journal* 103: 619–637.
Hinderer J, Legros H, and Amalvict M (1982) A search for the Chandler and nearly diurnal free wobbles using Liouville equations. *Geophysical Journal of Royal Astronomical Society* 71: 303–332.
Holin I (1999) Very long baseline radar interferometry for investigation of Mercury's spin and interior (in Russian). *Uspeckhi Sovremennoi Radioelectroniki* 7: 16–28.
Holme R (1998) Electromagnetic core-mantle coupling. I. Explaining decadal changes in the length of day. *Geophysical Journal International* 132: 167–180.
Hough SS (1895) The oscillations of a rotating ellipsoidal shell containing fluid. *Philosophical Transactions. Royal Society of London* A186: 469–506.
James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, USA: The University of Arizona Press.
Jault D, Gire D, and Le Mouel JL (1988) Westward drift, core motions and exchanges of angular momentum between core and mantle. *Nature* 333: 353–356.
Jeffreys H (1952) *The Earth its Origin, History, and Physical Constitution*. Cambridge: University Press.
Jehn R, Corral C, and Giampieri G (2004) Estimating Mercury's 88-day libration amplitude from orbit. *Planetary and Space Science* 52: 727–732.
Karatekin Ö, de Viron O, Lambert S, et al. (2011) Atmospheric angular momentum variations of Earth, Mars and Venus at seasonal timescales. *Planetary and Space Science* 59: 923–933.
Karatekin Ö, Van Hoolst T, and Dehant V (2006) Martian global scale CO_2 exchange from time-variable gravity measurements. *Journal of Geophysical Research* 111(E6). http://dx.doi.org/10.1029/2005JE002591, E06003.
Karatekin Ö, Van Hoolst T, Tastet J, de Viron O, and Dehant V (2006) The effects of seasonal mass redistribution and interior structure on Length-of-day variations of Mars. *Advances in Space Research* 38(4): 739–744. http://dx.doi.org/10.1016/j.asr.2005.03.117.
Kaula WM (1964) Tidal dissipation by solid friction and the resulting orbital motion. *Reviews of Geophysics* 2: 661–685.
Kaula WM (2000) *Theory of Satellite Geodesy*. New York, USA: Dover Publications.
Kiefer WS, Bills B, and Nerem RS (1996) An inversion of gravity and topography for mantle and crustal structure on Mars. *Journal of Geophysical Research* 101: 9239–9252.
Kieffer HH, Jakosky BM, and Snyder CW (1992) The planet Mars: From antiquity to the present. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1–33. Tucson, USA: The University of Arizona Press.
Kieffer HH and Titus TN (2001) TES mapping of the Mars north seasonal cap. *Icarus* 154: 162180. http://dx.doi.org/10.1006/icar.2001.6670.
Kieffer HH, Titus TN, Mullins KF, and Christensen PR (2000) Mars south polar spring and summer behavior observed by TES: Seasonal cap evolution controlled by frost grain size. *Journal of Geophysical Research* 105: 96539699.
Kinoshita H (1977) Theory of the rotation of the rigid Earth. *Celestial Mechanics* 15: 277–326.
Kokubo E and Ida S (2007) Formation of terrestrial planets from protoplanets II: Statistics of planetary spin. *Astrophysical Journal* 671: 2082–2090.
Koning A and Dumberry M (2013) Internal forcing of Mercury's long period free librations. *Icarus* 223(1): 40–47.
Konopliv AS, Asmar SW, Folkner WM, et al. (2011) Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters. *Icarus* 211(1): 401428. http://dx.doi.org/10.1016/j.icarus.2010.10.004.
Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.
Konopliv AS, Borderies NJ, Chodas PW, et al. (1993) Venus gravity and topography: 60th degree and order model. *Geophysical Research Letters* 20: 2403–2406.
Konopliv AS and Sjogren WL (1994) Venus spherical harmonic gravity model to degree and order 60. *Icarus* 112: 42–54.
Konopliv AS and Yoder CF (1996) Venus k_2 tidal Love number from Magellan and PVO tracking data. *Geophysical Research Letters* 23: 1857–1860.
Konopliv AS, Yoder CF, Standish EM, Yuan D-N, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 2350. http://dx.doi.org/10.1016/j.icarus.2005.12.025.
Kundt W (1977) Spin and atmospheric tides of Venus. *Astronomy and Astrophysics* 60: 85–91.
Lago B and Cazenave A (1979) Possible dynamical evolution of the rotation of Venus since formation. *Moon and the Planets* 21: 127–154.
Lambeck K (1980) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. Cambridge, UK: Cambridge University Press.
Laskar J (1994) Large scale chaos in the Solar System. *Astronomy and Astrophysics* 287: L9–L12.
Laskar J, Correia ACM, Gastineau M, Joutel F, Levrard B, and Robutel P (2004) Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* 170: 343–364.
Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.
Le Maistre S, Rosenblatt P, Rivoldini A, Dehant V, Marty J-C, and Karatekin Ö (2012) Lander radio science experiment with a direct link between Mars and the Earth. *Planetary and Space Science* 68: 105–122.
Lebonnois S, Hourdin F, Eymet V, Crespin A, Fournier R, and Forget F (2010) Superrotation of Venus' atmosphere analysed with a full General Circulation Model. *Journal of Geophysical Research (Planets)* 115: 6006.
Levrard B and Laskar J (2003) Climate friction and the Earth's obliquity. *Geophysical Journal International* 154: 970–990.
Lewis SR, Collins M, Read PL, et al. (1999) A climate database for Mars. *Journal of Geophysical Research* 104(E10): 24177–24194.
Lissauer JJ, Berman AF, Greenzweig Y, and Kary DM (1997) Accretion of mass and spin angular momentum by a planet on an eccentric orbit. *Icarus* 127: 65–92.
Lissauer JJ, Dones L, and Ohtsuki K (2000) Origin and evolution of terrestrial planet rotation. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 101–112. Tucson, USA: The University of Arizona Press.
Lissauer JJ and Kary DM (1991) The origin of the systematic component of planetary rotation. I. Planet on a circular orbit. *Icarus* 93: 126–159.
Litvak ML, Mitrofanov IG, Kozyrev AS, et al. (2004) Seasonal carbon dioxide depositions on the Martian surface as revealed from neutron measurements by the HEND Instrument Onboard the 2001 mars odyssey spacecraft. *Solar System Research* 38: 167–177.
Litvak ML, Nitrofanov IG, Sanin AB, et al. (2006) Comparison between polar regions of Mars from HEND/Odyssey data. *Icarus* 180: 23–37. http://dx.doi.org/10.1016/j.icarus.20005.08.009.
Lodders K and Fegley B Jr. (1998) *The Planetary Scientist's Companion*. New York, USA: Oxford University Press.
MacDonald GJF (1964) Tidal friction. *Reviews of Geophysics* 2: 467–541.
Manglik A, Wicht J, and Christensen UR (2010) A dynamo model with double diffusive convection for Mercury's core. *Earth and Planetary Science Letters* 289: 619628.
Margot J (2009) A Mercury orientation model including non-zero obliquity and librations. *Celestial Mechanics and Dynamical Astronomy* 105: 329336.
Margot J, Peale S, Jurgens RF, Slade MA, and Holin IV (2007) Large longitude libration of mercury reveals a molten core. *Science* 316: 710–714.
Margot J, Peale S, Solomon SC, et al. (2012) Mercury's moment of inertia from spin and gravity data. *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2012JE004161, E00L09.
Mathews PM, Buffett BA, Herring TA, and Shapiro II (1991) Forced nutations of the Earth: Influence of inner core dynamics 1. Theory. *Journal of Geophysical Research* 96(B5): 8219–8242.
Mathews PM and Guo JY (2005) Viscoelectromagnetic coupling in precession-nutation theory. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2003JB002915, B02402.

Mathews PM, Herring TA, and Buffett BA (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000390, ETG 3-1.

Matsuyama I and Manga M (2010) Mars without the equilibrium rotational figure, Tharsis, and the remnant rotational figure. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2010JE003686, E12020.

Matsuyama I, Mitrovica JX, Manga M, Perron JT, and Richards MA (2006) Rotational stability of dynamic planets with elastic lithospheres. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JE002447, E02003.

Matsuyama I and Nimmo F (2009) Gravity and tectonic patterns of Mercury: Effect of tidal deformation, spin–orbit resonance, nonzero eccentricity, despinning, and reorientation. *Journal of Geophysical Research* 114, E01010.

McCue J and Dormand JR (1993) Evolution of the spin of Venus. *Earth, Moon, and Planets* 63: 209–225.

McNamee JB, Borderies NJ, and Sjogren WL (1993) Venus: Global gravity and topography. *Journal of Geophysical Research* 98(E5): 9113–9128.

Melchior P (1983) *The Tides of the Planet Earth*. Oxford, UK: Pergamon Press.

Michel NC, Hauck SA II, Solomon SC, Phillips RJ, Roberts JH, and Zuber MT (2012) Implications of MESSENGER observations for mantle convection on Mercury. In: *43rd Lunar and Planetary Science Conference*, Abstract No. 1671.

Molodensky SM, Zharkov VN, and Gudkova TV (2009) On models of Mars' interior and amplitudes of forced nutations 2. The effects of liquid core and mantle elasticity. *Physics of the Earth and Planetary Interiors* 172: 335–344.

Moritz H and Mueller II (1987) *Earth Rotation. Theory and Observation*. New York, USA: Ungar.

Mottinger NA, Sjogren WL, and Bills BG (1985) Venus gravity: A harmonic analysis and geophysical implications. *Journal of Geophysical Research* 90: C739–C756.

Munk WH and MacDonald GJF (1960) *The Rotation of the Earth: A Geophysical Discussion*. Cambridge: Cambridge University Press.

Murray CD and Dermott SF (1999) *Solar System Dynamics*. Cambridge, UK: Cambridge University Press.

Nerem RS, Bills BG, and McNamee JB (1993) A high resolution gravity model for Venus: GVM-1. *Geophysical Research Letters* 20: 599–602.

Néron de Surgy O and Laskar J (1997) On the long term evolution of the spin of the Earth. *Astronomy and Astrophysics* 318: 975–989.

Ohtsuki K and Ida S (1998) Planetary rotation by accretion of planetesimals with uniform spatial distribution formed by the planet's gravitational perturbation. *Icarus* 131: 393–420.

Owen T, Biemann K, Rushneck DR, Biller JE, Howarth DW, and Lafleur AL (1977) The composition of the atmosphere at the surface of Mars. *Journal of Geophysical Research* 82: 4635–4639.

Pais MA, Le Mouel JL, Lambeck K, and Poirier JP (1999) Late Cambrian paradoxical glaciation and obliquity of the Earth—A discussion of dynamical constraints. *Earth and Planetary Science Letters* 174: 155–171.

Peale SJ (1969) Generalized Cassini laws. *Astronomical Journal* 74: 483–489.

Peale SJ (1972) Determination of parameters related to the interior of Mercury. *Icarus* 17: 168–173.

Peale SJ (1973) Some effects of elasticity on lunar rotation. *The Moon* 8: 515–531.

Peale SJ (1974) Possible histories of the obliquity of Mercury. *Astronomical Journal* 79: 722–744.

Peale SJ (1976a) Inferences from the dynamical history of Mercury's rotation. *Icarus* 28: 459–467.

Peale SJ (2005) The free precession and libration of Mercury. *Icarus* 178: 4–18.

Peale SJ (2006) The proximity of Mercury's spin to Cassini state 1 from adiabatic invariance. *Icarus* 181: 338–347.

Peale SJ and Boss AP (1977) A spin–orbit constraint on the viscosity of a Mercurian liquid core. *Journal of Geophysical Research* 82: 743–749.

Peale SJ, Margot JL, and Yseboodt M (2009) Resonant forcing of Mercury's libration in longitude. *Icarus* 199: 18.

Peale SJ, Phillips RJ, Solomon SC, Smith DE, and Zuber MT (2002) A procedure for determining the nature of Mercury's core. *Meteoritics and Planetary Science* 37: 1269–1283.

Peale SJ, Yseboodt M, and Margot JL (2007) Long-period forcing of Mercury's libration in longitude. *Icarus* 187: 365373.

Pettengill GH and Dyce RB (1965) A radar determination of the rotation of the planet Mercury. *Nature* 206: 1240.

Pfyffer G, Van hoolst T, and Dehant V (2011) Librations and obliquity of Mercury from the BepiColombo radio-science and camera experiments. *Planetary and Space Science* 59(9): 848–861.

Philip JR (1979) Angular momentum of seasonally condensing atmospheres, with reference to Mars. *Geophysical Research Letters* 6: 727–730.

Poincare H (1910) Sur la praécession des corps déformables. *Bulletin of the American Astronomical Society* 27: 321–356.

Pollack JB, Haberle RM, Schaeffer J, and Lee H (1990) Simulations of the general circulation of the Martian atmosphere: 1. Polar processes. *Journal of Geophysical Research* 95: 1447–1474.

Prettyman TH, Feldman WC, and Titus TN (2009) Characterization of Mars seasonal caps using neutron spectroscopy. *Journal of Geophysical Research* 114. http://dx.doi.org/10.1029/2008JE00327, E08005.

Rambaux N and Bois E (2004) Theory of the Mercury's spin–orbit motion and analysis of its main librations. *Astronomy and Astrophysics* 413: 381–393.

Rambaux N, Van Hoolst T, Dehant V, and Bois E (2007) Inertial core–mantle coupling and libration of Mercury. *Astronomy and Astrophysics* 468: 711–719. http://dx.doi.org/10.1051/0004-6361:20053974.

Reasenberg RD, Eubanks TM, MacNeil PE, and Shapiro II (1980) Mars rotation: Detection of seasonal rate variation. *Bulletin of the American Astronomical Society* 12: 721.

Reasenberg RD and King RW (1979) The rotation of Mars. *Journal of Geophysical Research* 84(B11): 6231–6240.

Reese CC and Peterson PE (2002) Thermal evolution of Mercury in the conductive regime and implications for magnetism. *Lunar and Planetary Science* XXXIII.

Rivoldini A and Van Hoolst T (2013) The interior structure of Mercury constrained by the low-degree gravity field and the rotation of Mercury. *Earth, Planets and Space* 377378: 6272.

Rivoldini A, Van Hoolst T, and Verhoeven O (2009) The interior structure of Mercury and its core sulfur content. *Icarus* 201: 12–30.

Rivoldini A, Van Hoolst T, Verhoeven O, Mocquet A, and Dehant V (2011) Geodesy constraints on the interior structure and composition of Mars. *Icarus* 213: 451–472. http://dx.doi.org/10.1016/j.icarus.2011.03.024.

Roberts PH and Stewartson K (1965) On the motion of a liquid in a spheroidal cavity of a precessing rigid body. II. *Proceedings of the Cambridge Philosophical Society* 61: 279–288.

Roberts JH and Zhong SJ (2004) Plume-induced topography and geoid anomalies and their implications for the Tharsis rise on Mars. *Journal of Geophysical Research* 109, E03009.

Rochester MG (1976) The secular decrease of obliquity due to dissipative core–mantle coupling. *Geophysical Journal of the Royal Astronomical Society* 46: 109–126.

Roosbeek F (1999) Analytical developments of rigid mars nutation and tide generating potential series. *Celestial Mechanics and Dynamical Astronomy* 75: 285–300.

Rubincam DP (1990) Mars: Change in axial tilt due to the climate? *Science* 248: 720–721.

Russell CT, Baker DN, and Slavin JA (1988) The magnetosphere of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AR, USA: Univ. Arizona Press.

Sanchez B, Haberle R, and Schaeffer J (2004) Atmospheric rotational effects on Mars based on the NASA Ames general circulation model: Angular momentum approach. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2004JE002254, E08005.

Sanchez BV, Rowlands DD, Haberele RM, and Schaeffer J (2003) Atmospheric rotational effects on Mars based on the NASA Ames general circulation model. *Journal of Geophysical Research* 108(E5): 5040. http://dx.doi.org/10.1029/2002JE001984.

Sasao T, Oklamoto I, and Sakai S (1977) Dissipative core–mantle coupling and nutational motion of the Earth. *Publications of the Astronomical Society of Japan* 29: 83–105.

Sasao T, Okubo S, and Saito M (1980) A simple theory on the dynamical effects of a stratified fluid core upon nutational motion of the Earth. In: Fedorov EP, Smith ML, and Bender PL (eds.) *Proceedings of IQU Symposium 78: Nutation and the Earth's rotation. Kiev, May 1977. Reidel, Dordrecht* 165–183.

Schlichting HE and Sari R (2007) The effect of semicollisional accretion on planetary spins. *The astrophysical Journal* 658: 593–597.

Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 429–460. Tucson, AR, USA: Univ. Arizona Press.

Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, USA: The University of Arizona Press.

Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95(B9): 14095–14104.

Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge, UK: Cambridge University Press.

Seidelmann PK (2005) *Explanatory Supplement to the Astronomical Almanac*. California: University Science Books.

Shapiro II, Chandler JF, Campbell DB, Hine AA, and Stacy NJS (1990) The spin vector of Venus. *Astronomical Journal* 100: 1363–1368.
Shen M and Zhang CZ (1989) Dynamical evolution of the rotation of Venus. *Earth, Moon, and Planets* 43: 275–287.
Smart WM (1953) *Celestial Mechanics*. London: Longmans, Green and Co.
Smith ML (1977) Wobble and nutation of the Earth. *Geophysical Journal of Royal Astronomical Society* 50: 103–140.
Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
Smith ML and Dahlen FA (1981) The period and Q of the Chandler wobble. *Geophysical Journal* 64: 223–281.
Smith DE, Zuber MT, Haberle RM, Rowlands DD, and Murphy JR (1999) The Mars seasonal CO_2 cycle and the time variation of the gravity field: A general circulation model simulation. *Journal of Geophysical Research* 104: 1885–1896.
Smith DE, Zuber MT, and Neumann GA (2001) Seasonal variations of snow depth on Mars. *Science* 294: 2141–2146.
Smith DE, Zuber MT, Phillips RJ, et al. (2012) Gravity field and internal structure of mercury from MESSENGER. *Science* 336: 214.
Sohl F and Spohn T (1997) The structure of Mars: Implications from SNC-meteorites. *Journal of Geophysical Research* 102: 1613–1635.
Spada G and Alfonsi L (1998) Obliquity variations due to climate friction on Mars: Darwin versus layered models. *Journal of Geophysical Research* 103(E12): 28599–28605.
Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–237.
Spohn T, Sohl F, Wieczerkowski K, and Conzelmann V (2001) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary and Space Science* 49: 1561–1570.
Stacey FD (1992) *Physics of the Earth*. Brisbane, Australia: Brookfield Press.
Stanley S, Bloxham J, Hutchinson WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.
Stevenson DJ (2001) Mars core and magnetism. *Nature* 421: 214–219.
Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
Stewartson K and Roberts PH (1963) On the motion of a liquid in a spheroidal cavity of a precessing rigid body. *Journal of Fluid Mechanics* 17: 1–20.
Takahashi F and Matsushima M (2006) Dipolar and non-dipolar dynamos in a thin shell geometry with implications for the magnetic field of Mercury. *Geophysical Research Letters* L10202. http://dx.doi.org/10.1029/2006GL025792.
Thompson W (1882) On the thermodynamic acceleration of the earth's rotation. *Proceedings of Royal Society of Edinburgh* 11: 396–405.
Thomson DJ (1990) Quadratic inverse spectrum estimates: Applications to paleoclimatology. *Philosophical Transactions of Royal Society of London* A332: 539–597.
Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
Van den Acker E, Van Hoolst T, de Viron O, et al. (2002) Influence of the winds and the CO_2 mass exchange between the atmosphere and the polar ice caps on Mars' orientation parameters. *Journal of Geophysical Research* http://dx.doi.org/10.1029/2000JE001539.
Van Hoolst T and Dehant V (2002) Influence of triaxiality and second-order terms in flattenings on the rotation of terrestrial planets: I. Formalism and rotational normal modes. *Physics of the Earth and Planetary Interiors* 134: 17–33.
Van Hoolst T, Dehant V, and Defraigne P (2000a) Sensitivity of the Free Core Nutation and the Chandler Wobble to changes in the interior structure of Mars. *Physics of the Earth and Planetary Interiors* 117: 397–405.
Van Hoolst T, Dehant V, and Defraigne P (2000b) Chandler wobble and free core nutation for Mars. *Planetary and Space Science* 48: 1145–1151.
Van Hoolst T, Dehant V, Roosbeek F, and Lognonn P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 161(2): 281–296.
Van Hoolst T and Jacobs C (2003) Mercury's tides and interior structure. *Journal of Geophysical Research* 108(E11): 5121. http://dx.doi.org/10.1029/2003JE002126.
Van Hoolst T, Rivoldini A, Baland R-M, and Yseboodt M (2012) The effect of tides and an inner core on the forced 88 day longitudinal libration of Mercury. *Earth and Planetary Science Letters* 333–334: 83–90.
Van Hoolst T, Sohl F, Holin I, Verhoeven O, Dehant V, and Spohn T (2007) Mercury's interior structure, rotation, and tides. *Space Science Reviews* 132(2–4): 203–227. http://dx.doi.org/10.1007/s11214-007-9202-6.
Vilim R, Stanley S, and Hauck SA II (2010) Iron snow zones as a mechanism for generating Mercury's weak observed magnetic field. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2009JE0035280, E11003.
Wahr JM (1981a) A normal mode expansion for the forced response of a rotating earth. *Geophysical Journal of the Royal Astronomical Society* 64: 651–675.
Wahr JM (1981b) The forced nutations of an elliptical, rotating, elastic and oceanless earth. *Geophysical Journal of the Royal Astronomical Society* 64: 705–727.
Ward WR (1974) Climatic variations on Mars. 1. Astronomical theory of insulation. *Journal of Geophysical Research* 79: 3375–3386.
Ward WR (1975) Tidal friction and generalized Cassini's laws in the solar system. *Astronomical Journal* 80: 64–70.
Ward WR (1992) Long-term orbital dynamics and spin dynamics of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 298–320. Tucson, USA: The University of Arizona Press.
Wieczorek MA, Correia ACM, Le Feuvre M, Laskar J, and Rambaux N (2011) Mercury's spin–orbit resonance explained by initial retrograde and subsequent synchronous rotation. *Nature Geosciences* 5: 18.
Willemann RJ (1984) Reorientation of planets with elastic lithospheres. *Icarus* 60: 701709.
Williams DM, Kasting JF, and Frakes LA (1998) Low-latitude glaciation and rapid changes in the Earth's obliquity explained by obliquity-oblateness feedback. *Nature* 396: 453–455.
Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
Yoder CF (1995) Venus' free obliquity. *Icarus* 117: 250–286.
Yoder CF (1997) Venusian spin dynamics. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1087–1124. Tucson: University of Arizona Press.
Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303. http://dx.doi.org/10.1126/science.1079645.
Yoder CF and Standish EM (1997) Martian precession and rotation from Viking lander range data. *Journal of Geophysical Research* 102(E2): 4065–4080.
Yoder CF and Ward WR (1979) Does Venus wobble? *Astrophysical Journal* 233: L33–L37.
Yoder CF, Williams JG, and Parke ME (1981) Tidal variations of earth rotation. *Journal of Geophysical Research* 86: 881–891.
Yseboodt M, Barriot J-P, and Dehant V (2003) Analytical modeling of the Doppler tracking between a lander and a Mars orbiter in terms of rotational dynamics. *Journal of Geophysical Research* 108(E7): 5075. http://dx.doi.org/10.1029/2003JE002045.
Yseboodt M and Margot J-L (2006) Evolution of Mercury's obliquity. *Icarus* 181: 327–337.
Yseboodt M, Margot J-L, and Peale SJ (2010) Analytical model of the long-period forced longitude librations of Mercury. *Icarus* 207: 536–544.
Yseboodt M, Rivoldini A, Van Hoolst T, and Dumberry M (2013) Influence of an inner core on the long-period forced librations of Mercury. *Icarus* 226: 4151.
Zharkov VN and Gudkova TV (2005) Construction of Martian interior model. *Solar System Research* 39: 343–373.
Zhong SJ (2009) Migration of Tharsis volcanism on Mars caused by differential rotation of lithosphere. *Nature Geosciences* 2: 1923.

This page intentionally left blank

10.05 Gravity and Topography of the Terrestrial Planets

MA Wieczorek, University Paris Diderot, Paris Cedex 13, France

© 2015 Elsevier B.V. All rights reserved.

10.05.1	**Introduction**	153
10.05.2	**Mathematical Preliminaries**	154
10.05.2.1	Spherical Harmonics	155
10.05.2.2	Gravity, Potential, and Geoid	156
10.05.3	**The Data**	158
10.05.3.1	Earth	158
10.05.3.1.1	Topography	158
10.05.3.1.2	Gravity	159
10.05.3.1.3	Spectral analysis	159
10.05.3.2	Venus	161
10.05.3.2.1	Topography	161
10.05.3.2.2	Gravity	162
10.05.3.2.3	Spectral analysis	162
10.05.3.3	Mars	164
10.05.3.3.1	Topography	164
10.05.3.3.2	Gravity	166
10.05.3.3.3	Spectral analysis	166
10.05.3.4	Mercury	167
10.05.3.4.1	Topography	167
10.05.3.4.2	Gravity	167
10.05.3.4.3	Spectral analysis	169
10.05.3.5	The Moon	169
10.05.3.5.1	Topography	169
10.05.3.5.2	Gravity	170
10.05.3.5.3	Spectral analysis	170
10.05.4	**Methods for Calculating Gravity from Topography**	172
10.05.5	**Crustal Thickness Modeling**	173
10.05.6	**Admittance Modeling**	174
10.05.6.1	Spatial Domain	174
10.05.6.2	Spectral Domain	177
10.05.7	**Localized Spectral Analysis**	179
10.05.7.1	Multitaper Spectral Analysis	179
10.05.7.2	Wavelet Analysis	180
10.05.8	**Summary of Major Results**	181
10.05.8.1	Earth	181
10.05.8.2	Venus	182
10.05.8.3	Mars	183
10.05.8.4	Mercury	184
10.05.8.5	The Moon	185
10.05.9	**Future Developments and Concluding Remarks**	186
Acknowledgments		187
References		187

10.05.1 Introduction

One of the most precise ways to investigate the subsurface structure of a planet is through the analysis of seismic waves. While such endeavors have proved to be remarkably successful for Earth, the emplacement of a geophysical network that includes seismometers on a terrestrial body such as Mercury, Venus, Mars, or the Moon is both costly and technologically challenging (see Chapter 10.03). Fortunately, alternative means exist to probe the interior structure of the terrestrial planets from orbit. One such method, reviewed in Chapter 10.04, is through analyses of a planet's time-variable rotation. By the measurement of precessional and tidal signals, the moment of inertia factor and k_2 Love number can be constrained, from which constraints on the density and viscosity profile can be obtained. The magnetic induction response of a planet to time-variable magnetic fields can be used to determine how the body's electrical conductivity (and hence

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00169-X

composition) varies with depth. Finally, because the gravitational field of a planet is sensitive to its internal density structure, another powerful method that can be used to probe the subsurface, and which is the subject of this review, is the joint analysis of gravitational and topographic data.

The past two decades have witnessed dramatic advancements in our understanding and knowledge of the gravitational fields and topography of the terrestrial planets. These advances are intimately related to the acquisition of radio tracking data from orbiting spacecraft, measurements of the range and range rate between co-orbiting satellites, measurements of gravitational gradients from orbit, the collection of data from orbiting altimeters, and the construction of terrain models from stereo imagery. As examples, from data collected by NASA's Lunar Reconnaissance Orbiter (LRO; Vondrak et al., 2010) and Gravity Recovery and Interior Laboratory (GRAIL) (Zuber et al., 2013a) missions, respectively, launched in 2009 and 2011, the gravity and topography of the Moon are now known better globally than any other planetary object, including Earth. The topography of Venus was mapped by the technique of radar altimetry during the Magellan mission (placed in orbit in 1990; Saunders et al., 1990), and radio tracking data acquired during this mission has led to the construction of a global gravity model. From the laser altimeter onboard the Mars Global Surveyor (MGS) mission (placed in orbit in 1997; Smith et al., 2001a), the topography of Mars has been mapped to high resolution, and radio tracking of this and subsequent spacecraft has allowed for continued improvement to the Martian gravity field. In a similar manner, radio tracking and laser altimetry are being used by the MESSENGER spacecraft that is currently in orbit about Mercury (Solomon et al., 2001). Finally, the topography and gravitational field of Earth are continually being refined, most notably by a series of recent and ongoing missions such as GRACE (Gravity Recovery and Climate Experiment; Tapley et al., 2005) and GOCE (Gravity Field and Steady-State Ocean Circulation Explorer; Drinkwater et al., 2003).

From a geophysical and comparative planetology perspective, the motivation for obtaining high-resolution gravitational and topographic data is to place constraints on the formation, interior structure, and geologic evolution of a planet. The magnitude and direction of the gravitational acceleration exterior to a planet are completely determined by its internal distribution of mass, a result of Newton's law of gravitation. When combined with topographic data and reasonable geologic assumptions, it becomes possible to invert for important geophysical parameters, such as crustal thickness, elastic thickness, and crustal and mantle densities. These model parameters can then be used to address questions concerning planetary differentiation, crust formation, thermal evolution, and magmatic processes. Since the spatial resolving power of gravitational measurements decreases with increasing distance from the object, these investigations are generally restricted to the crust and upper mantle.

Very few research articles have been written that review the gravitational fields and topography of the terrestrial planets from a comparative planetology perspective: useful summaries can be found in Phillips and Lambeck (1980), Balmino (1993), Bills and Lemoine (1995), and Rummel (2005). The purpose of this chapter is both to review the current state of knowledge of the gravitational fields and topography of the terrestrial planets and to review the tools that are used to describe and analyze these data. While gravitational and topographic datasets can each be used independently to make inferences about the interior structure of a planet, such results are often based upon hypotheses that are not easily testable or models that are highly underconstrained. Thus, although regional topographic models have been constructed for some moons and asteroids (such as Phobos (Wählisch et al., 2010), 433 Eros (Gaskell et al., 2008; Zuber et al., 2000a), Itokawa (Abe et al., 2006; Gaskell et al., 2008), Vesta (Jaumann et al., 2012), Ganymede (Giese et al., 1998), Europa (Nimmo et al., 2003a,b; Nimmo et al., 2007), Iapetus (Giese et al., 2008), Mimas (Dermott and Thomas, 1988), Enceladus (Thomas et al., 2007), Tethys (Thomas et al., 2007), Dione (Thomas et al., 2007), Rhea (Thomas et al., 2007), and many of the irregularly shaped satellites of Saturn (Thomas, 2010)) and the longest wavelength gravitational fields and topography have been constrained for others (such as Io (Anderson et al., 2001a; Thomas et al., 1998), Europa (Anderson et al., 1998), Ganymede (Anderson et al., 1996b; Palguta et al., 2006), Callisto (Anderson et al., 2001b), Rhea (Anderson and Schubert, 2010), Titan (Iess et al., 2010; Lorenz et al., 2013), and Vesta (Konopliv et al., 2013a)), this chapter will focus on those planetary bodies for which the gravity and topography are both characterized to high resolution, namely, Earth, Venus, Mars, Mercury, and the Moon.

The outline of this chapter is as follows. First, in **Section 10.05.2**, a general review of the mathematical formalism that is used in describing the properties of gravitational fields and topography is given. In **Section 10.05.3**, the basic properties of the gravitational fields and topography of Earth, Venus, Mars, Mercury, and the Moon are characterized. Following this, **Sections 10.05.4–10.05.7** discuss the relationship between gravity and topography and how the two datasets can be used to invert for geophysical parameters. These methods include crustal thickness modeling, the analysis of geoid/topography ratios (GTRs), modeling of the spectral admittance and correlation functions, and localized spectral analysis and wavelet techniques. **Section 10.05.8** summarizes the major results that have been obtained for these planetary bodies, and **Section 10.05.9** concludes by discussing future developments that can be expected in this domain.

10.05.2 Mathematical Preliminaries

This section reviews the mathematical background that is required to understand how gravitational fields and topography are characterized and manipulated. As most analysis techniques make use of spherical harmonics, the first subsection defines these functions and introduces certain concepts such as the power spectrum. The following subsection gives definitions that are specific to the gravitational potential, gravitational field, and geoid. For further details, the reader is referred to several key books and articles such as **Chapter 3.02**, Jeffreys (1976), Kaula (1967), Kaula (2000), Heiskanen and Moritz (1967), Lambeck (1988), and Hofmann-Wellenhof and Moritz (2006).

10.05.2.1 Spherical Harmonics

Spherical harmonics are the natural set of orthogonal basis functions on the surface of the sphere. As such, any real square-integrable function can be expressed as a linear combination of these as

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\theta, \phi) \quad [1]$$

where Y_{lm} is the spherical harmonic function of degree l and order m, f_{lm} is the corresponding expansion coefficient, and θ and ϕ represent the position on the sphere in terms of planetocentric colatitude and longitude, respectively. Alternative representations include the use of ellipsoidal harmonics (e.g., Garmier and Barriot, 2001; Jekeli, 1988) when the function is known globally and Slepian functions (Simons et al., 2006) or spherical cap harmonics (e.g., Haines, 1985; Hwang and Chen, 1997; Thébault et al., 2006, 2004) when the function is known only regionally. This chapter will use exclusively planetocentric coordinates, where colatitude is measured with respect to the polar axis, but the reader should be aware that many datasets for Earth make use of planetographic coordinates. Planetographic latitude differs from planetocentric latitude in that it is defined as the angle between the normal to an ellipsoid and the equatorial plane.

In geodesic and gravitational applications, the real spherical harmonics are defined as

$$Y_{lm}(\theta, \phi) = \begin{cases} \overline{P}_{lm}(\cos\theta)\cos m\phi & \text{if } m \geq 0 \\ \overline{P}_{l|m|}(\cos\theta)\sin|m|\phi & \text{if } m < 0 \end{cases} \quad [2]$$

where the normalized associated Legendre functions are given by

$$\overline{P}_{lm}(\mu) = \sqrt{(2-\delta_{0m})(2l+1)\frac{(l-m)!}{(l+m)!}} P_{lm}(\mu) \quad [3]$$

and where δ_{ij} is the Kronecker delta function. The unnormalized Legendre functions in the aforementioned equation are defined in relation to the Legendre polynomials by

$$P_{lm}(\mu) = \left(1-\mu^2\right)^{m/2} \frac{\mathrm{d}^m}{\mathrm{d}\mu^m} P_l(\mu) \quad [4]$$

$$P_l(\mu) = \frac{1}{2^l l!} \frac{\mathrm{d}^l}{\mathrm{d}\mu^l} \left(\mu^2 - 1\right)^l \quad [5]$$

The normalized associated Legendre functions are orthogonal for a given value of m according to

$$\int_{-1}^{1} \overline{P}_{lm}(\mu)\overline{P}_{l'm}(\mu)\mathrm{d}\mu = 2(2-\delta_{0m})\delta_{ll'} \quad [6]$$

and the spherical harmonics are orthogonal over both l and m with the normalization

$$\int_{\Omega} Y_{lm}(\theta, \phi) Y_{l'm'}(\theta, \phi)\mathrm{d}\Omega = 4\pi\delta_{ll'}\delta_{mm'} \quad [7]$$

where $\mathrm{d}\Omega = \sin\theta\ \mathrm{d}\theta\ \mathrm{d}\varphi$. Multiplying eqn [1] by $Y_{l'm'}$ and integrating over all space, the spherical harmonic coefficients of the function f can be obtained from the integral

$$f_{lm} = \frac{1}{4\pi} \int_{\Omega} f(\theta, \phi) Y_{lm}(\theta, \phi)\mathrm{d}\Omega \quad [8]$$

A useful visualization property of the spherical harmonic functions is that they possess $2|m|$ zero crossings in the longitudinal direction and $l-|m|$ zero crossings in the latitudinal direction. In addition, for a given spherical harmonic degree l, the equivalent Cartesian wavelength is $\lambda \approx 2\pi R/\sqrt{l(l+1)}$, where R is the mean planetary radius, a result known as the Jeans relation. It should be noted that those coefficients and spherical harmonics where $m=0$ are referred to as zonal, those with $l=|m|$ are sectoral, and the rest are tesseral. Furthermore, the zonal coefficients C_{l0} are often written as $-J_l$.

Using the orthogonality properties of the spherical harmonic functions, it is straightforward to verify that the total power of a function f is related to its spectral coefficients by a generalization of Parseval's theorem:

$$\frac{1}{4\pi} \int_{\Omega} f^2(\theta, \phi)\mathrm{d}\Omega = \sum_{l=0}^{\infty} S_{ff}(l) \quad [9]$$

where

$$S_{ff}(l) = \sum_{m=-l}^{l} f_{lm}^2 \quad [10]$$

is referred to as the power spectrum of the function. (The term 'power' is here used in the sense of the signal processing literature; namely, the square of the function divided by the area that it spans.) It can be shown that S_{ff} is unmodified by a rotation of the coordinate system. Similarly, the cross power of two functions f and g is given by

$$\frac{1}{4\pi} \int_{\Omega} f(\theta, \phi) g(\theta, \phi)\mathrm{d}\Omega = \sum_{l=0}^{\infty} S_{fg}(l) \quad [11]$$

where

$$S_{fg}(l) = \sum_{m=-l}^{l} f_{lm} g_{lm} \quad [12]$$

is defined as the cross power spectrum. If the functions f and g have a zero mean (i.e., their degree-0 terms are equal to zero), then $S_{ff}(l)$ and $S_{fg}(l)$ represent the contributions to the variance and covariance, respectively, for degree l. Some authors plot routinely the power per degree, $S/(2l+1)$, or the root-mean-square (rms) amplitude of the spherical harmonic coefficients, $\sqrt{S/(2l+1)}$.

One source of confusion with spherical harmonic analyses is that not all authors use the same definitions for the spherical harmonic and Legendre functions. In contrast to the 'geodesy' or '4π' normalization of eqn [7] (cf. Kaula, 2000), the seismology (e.g., Dahlen and Tromp, 1998) and physics (e.g., Varshalovich et al., 1988) communities often use orthonormal harmonics, whose squared integral is equal to unity. The geomagnetic community employs Schmidt seminormalized harmonics whose squared integral is $4\pi/(2l+1)$ (e.g., Blakely, 1995). A more subtle problem is related to a phase factor of $(-1)^m$ (the Condon–Shortley phase) that is sometimes applied to either eqn [3] or [4]. Whereas the spherical harmonics used by the geodesy and geomagnetic communities both exclude this phase factor, those in the physics and seismology communities often include it. To convert the spherical harmonic coefficients from one normalization to another (here labeled by

superscripts 1 and 2), it is necessary only to recognize that the product of the spherical harmonic coefficient and function is independent of the normalization:

$$f_{lm}^{(1)}Y_{lm}^{(1)} = f_{lm}^{(2)}Y_{lm}^{(2)} \qquad [13]$$

In order to obtain the spherical harmonic coefficients f_{lm} of a function f, it is necessary to be able to calculate accurately both the normalized Legendre functions of eqn [3] and the integral of eqn [8]. Methods for efficient calculation of the normalized associated Legendre functions depend upon the use of well-known three-term recursion formulas. If starting values used in the recursion are appropriately scaled, as is summarized by Holmes and Featherstone (2002), these can be computed to high accuracy up to a maximum spherical harmonic degree of about 2700. To obtain a similar accuracy at higher degrees would require the use of an alternative scaling algorithm (e.g., Fukushima, 2011; Jekeli et al., 2007).

The integrals of eqn [8] are most easily performed if the function f is known on a set of evenly spaced grid points in longitude. Numerical methods for calculating this integral involve Fourier transforming each latitudinal band and then integrating over latitude for each l and m (e.g., Sneeuw, 1994). If the function is sampled on an $n \times n$ grid, with n even, and if the function is known to be bandlimited to a maximum degree $n/2-1$, then the spherical harmonic transform can be computed exactly (see Driscoll and Healy, 1994). Alternatively, the integral over latitude can be performed using Gauss–Legendre quadrature (e.g., Press et al., 1992, pp. 140–149). While the integrand of eqn [8] is not in general a terminating polynomial, if the function is bandlimited to a maximum degree n, $n+1$ points in latitude are sufficient to calculate accurately the spherical harmonic expansion coefficients. Publicly available software packages, such as SHTOOLS developed by the author, are readily available for performing the calculations in this manuscript (see **Table 1** for a list of Internet resources).

10.05.2.2 Gravity, Potential, and Geoid

The gravitational acceleration $\mathbf{g}$ at position $\mathbf{r}$ that results from a mass dm at position $\mathbf{r}'$ is given by Newton's law of gravitation:

$$\mathbf{g}(\mathbf{r}) = -G\mathrm{d}m\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3} \qquad [14]$$

Given that the curl of $\mathbf{g}$ is zero, this vector force field is conservative, and the gravitational acceleration can be expressed as the gradient of a scalar potential U:

$$\mathbf{g}(\mathbf{r}) = \nabla U(\mathbf{r}) \qquad [15]$$

where the gradient operator in spherical coordinates is

$$\nabla = \hat{r}\frac{\partial}{\partial r} + \hat{\theta}\frac{1}{r}\frac{\partial}{\partial \theta} + \hat{\phi}\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi} \qquad [16]$$

The gravitational potential can be calculated at an arbitrary point by a simple integral over the mass distribution

$$U(\mathbf{r}) = \int_V \frac{G\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}V' \qquad [17]$$

where ρ is the mass density and V signifies that space occupied by the body. Whereas the sign convention of eqns [15] and [17] are consistent with that used in the geodesy literature, it should be noted that other disciplines, such as physics, place a negative sign in front of each of these equations.

Exterior to the mass distribution V, it can be shown that the potential satisfies Laplace's equation (e.g., Kaula, 2000, chap. 1):

$$\nabla^2 U(\mathbf{r}) = 0 \qquad [18]$$

By using this relationship and the method of separation of variables, the potential U exterior to V can be expressed alternatively as a sum of spherical harmonic functions:

$$U(\mathbf{r}) = \frac{GM}{r}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\frac{R_0}{r}\right)^l C_{lm}Y_{lm}(\theta,\phi) \qquad [19]$$

Here, the C_{lm}s represent the spherical harmonic coefficients of the gravitational potential at a reference radius R_0, G is the gravitational constant, and M is the total mass of the object. In practice, the infinite sum is truncated beyond a maximum degree L that is justified by the data resolution. The coefficients C_{lm} of eqn [19] are related uniquely to the internal mass distribution of the body, and methods for calculating these

Table 1 Internet resources

Resource	*Internet address*
SHTOOLS: Fortran 95 spherical harmonics software archive	http://shtools.ipgp.fr/
ETOPO1: Earth 1 arc minute topography model	http://www.ngdc.noaa.gov/mgg/global/global.html
GTOPO30: Earth 30 arc second topography model	https://lta.cr.usgs.gov/GTOPO30
GLOBE: Earth 30 arc second global topography	http://www.ngdc.noaa.gov/mgg/topo/globe.html
SRTM: Earth topography model	http://dds.cr.usgs.gov/srtm/
STRM30_PLUS: Earth 30 arc second topography and bathymetry model	http://topex.ucsd.edu/WWW_html/srtm30_plus.html
Earth2012: Spherical harmonic models of the Earth's topography and potential	http://geodesy.curtin.edu.au/research/models/Earth2012/
WGS84 ellipsoid and 'WGS84 EGM96 geoid'	http://earth-info.nga.mil/GandG/wgs84/
ICGEM: International Center for Global Earth Models (collection of gravity models)	http://icgem.gfz-potsdam.de/ICGEM/ICGEM.html
Planetary Data System (PDS) Geosciences node	http://geo.pds.nasa.gov/
Spherical harmonic models of planetary topography	http://www.ipgp.fr/~wieczor/SH/SH.html
GTDR3.2: Venusian topography	ftp://voir.mit.edu/pub/mg3003/
Generic Mapping Tools	http://gmt.soest.hawaii.edu/

are deferred until Section 10.05.4. Here, it is only noted that the degree-0 coefficient C_{00} is equal to unity and that if the coordinate system has been chosen such that it coincides with the center of mass of the object, the degree-1 terms (C_{10}, C_{11}, and $C_{1,-1}$) are identically zero. The negative order coefficients $C_{l,-m}$ are often referred to as S_{lm}.

As a result of the factor $(R_0/r)^l$ that multiplies each term in eqn [19], the high-frequency components in this sum (i.e., those with large l) will be attenuated when the observation radius r is greater than the reference radius R_0. In contrast, if the coefficients were determined at the altitude of an orbiting satellite, and if this equation were used to determine the potential field on the surface of the planet, then the high-frequency terms would instead be relatively amplified. Since the spherical harmonic coefficients always possess some uncertainty, which generally increases with increasing l, the process of downward continuing a potential field is not stable and must generally be compensated by some form of filtering (e.g., Fedi and Florio, 2002; Phipps Morgan and Blackman, 1993; Wieczorek and Phillips, 1998).

If the body under consideration is rotating, an additional nongravitational force must be taken into account when one is on the surface. In the reference frame of the rotating body, this can be accounted for by adding to eqn [19] a pseudopotential term that is a result of the centrifugal force. This rotational potential, as well as its spherical harmonic expansion, is given by

$$U^{\text{rot}} = \frac{\omega^2 r^2 \sin^2\theta}{2} = \omega^2 r^2 \left(\frac{1}{3} Y_{00} - \frac{1}{3\sqrt{5}} Y_{20} \right) \qquad [20]$$

where ω is the angular velocity of the rotating object. For some applications, especially concerning Earth and Moon, it is necessary to include the tidal potential of the satellite or parent body when calculating the potential (see Dermott, 1979; Zharkov et al., 1985). For a synchronously locked satellite on a circular orbit, the combined tidal and rotational potential in the rotating frame is given approximately by

$$U^{\text{tide+rot}} \simeq \omega^2 r^2 \left(\frac{1}{3} Y_{00} - \frac{5}{6\sqrt{5}} Y_{20} + \frac{1}{4} \sqrt{\frac{12}{5}} Y_{22} \right) \qquad [21]$$

For eccentric and inclined orbits, the reader is referred to Murray and Dermott (2008).

An important quantity in both geodesy and geophysics is the geoid, which is defined to be a surface that possesses a specific value of the potential. (Although one could come up with imaginative names for equipotential surfaces on Mercury, Venus, Eros, Io, and Titan, among others, the term geoid will here be used for all.) As there are no tangential forces on an equipotential surface (see eqn [15]), a static fluid would naturally conform to the geoid. The oceans of Earth are approximately static and are well approximated by such a surface. In geophysics, stresses within the lithosphere are often calculated by referencing the vertical position of a density contrast to an equipotential surface. This is necessary when performing lithospheric flexure calculations, especially when considering the longest wavelengths.

The height N of an equipotential surface above a spherical reference radius R can be obtained by approximating the potential by a Taylor series

$$U(R+N) \simeq U(R) + \frac{\mathrm{d}U(R)}{\mathrm{d}r} N + \frac{1}{2} \frac{\mathrm{d}^2 U(R)}{\mathrm{d}^2 r} N^2 \qquad [22]$$

and setting this expression equal to a constant

$$U(R+N) = \frac{GM}{R} + \frac{\omega^2 R^2}{3} \qquad [23]$$

the value of which is here chosen to be the degree-0 term of the potential at the reference radius R for a rotating planet. Since this equation is quadratic in N, the geoid height can be solved for analytically at any given position. Analytic expressions for the partial derivatives of the potential are obtained easily in the spectral domain from eqns [19] and [20] (see also eqn [25]). It is noted that an analytic expression for the geoid using a third-order expansion of eqn [22] also exists.

For some cases, it is sufficient to use only the first-order term of eqn [22]. If the first derivative of U is approximated by $-GM/R^2$, then the geoid is simply given by

$$N \simeq R \sum_{l=2}^{\infty} \sum_{m=-l}^{l} \left(\frac{R_0}{R} \right)^l C_{lm} Y_{lm} - \frac{\omega^2 R^4}{3\sqrt{5} GM} Y_{20} \qquad [24]$$

where the degree-1 terms have been assumed to be zero. Although the maximum difference between the first- and second-order methods is less than half a meter for the Moon, Mercury, and Venus, differences of up to 26 and 42 m are obtained for Earth and Mars, respectively. The poor performance of the first-order approximation for Earth and Mars is a result of the large rotationally induced flattening of these two planets.

Despite the simplicity of the aforementioned method for obtaining the height to an equipotential surface, the question arises as to *which* equipotential surface should be used. For Earth, a natural choice is the potential corresponding to mean sea level. However, for the other planets, the choice is more arbitrary (but often not important). As the aforementioned equations for calculating the potential are strictly valid only when the observation point is exterior to the body, one manner of picking a specific potential might be to choose that value for which all points on the geoid are exterior to the body. (Calculation of the potential below the surface of a planet would require knowledge of the interior density distribution.) Another approach that could be used would be to use the mean potential on the planet's equator. Alternatively, the geoid could be chosen such that it best fits, in a least squares sense, the observed shape of the planet (Ardalan et al., 2010).

The radial component of the gravitational field is calculated by taking the first radial derivative of eqn [19]. Ignoring the rotational potential, and using the sign convention that gravitational accelerations are positive when directed downward, this is given by the expression

$$g_r = \frac{GM}{r^2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left(\frac{R_0}{r} \right)^l (l+1) C_{lm} Y_{lm} \qquad [25]$$

Note that this equation differs from the potential only by the inclusion of the additional factors $1/r$ and $(l+1)$. The second factor gives a greater importance to the higher-degree terms, and it is for this reason that plots of the potential and geoid appear to be smooth when compared to components of the gravitational field. In terrestrial applications, it is common to calculate the gravitational field on the geoid. By inserting

eqn [24] into eqn [25] and ignoring rotational and higher-order terms, it can be shown that the radial component of the gravitational field on the geoid can be approximated simply by replacing the term $(l+1)$ in eqn [25] by $(l-1)$. If it is necessary to calculate the gravity field on an irregular surface, this can be accomplished by using a Taylor series that makes use of the known field and higher-order derivatives on a spherical interface (Hirt, 2012). The standard unit for quantifying gravitational perturbations is the galileo, where $1\ \text{Gal} = 10^{-2}\ \text{m s}^{-2}$, and when plotting gravitational anomalies in map form, it is conventional to use milligals.

Finally, it will be useful to characterize the relationship between the gravitational field and topography in the spectral domain. For this, let us presume that the radial gravity g_{lm} and topography h_{lm} are related by

$$g_{lm} = Q_{lm} h_{lm} + I_{lm} \quad [26]$$

where Q_{lm} is a linear nonisotropic transfer function and I_{lm} is that portion of the gravitational signal, such as noise, that is not predicted by the model (the topography can often be considered noise-free). If I_{lm} is uncorrelated with the topography, then it is straightforward to show that the expectation of S_{gh} will be unaffected by this signal. (This can be shown by multiplying eqn [26] by h_{lm}, summing over all m, and taking the expectation.) In contrast, the expectation of the gravitational power spectrum will depend upon the magnitude of I_{lm}. In particular, for the case where I_{lm} is a random variable that is independent of the surface topography, the expectation of the observed gravitational power spectrum is simply

$$S_{gg}^{\text{obs}}(l) = S_{gg}(l) + S_{II}(l) \quad [27]$$

where S_{gg} is the power spectrum predicted exclusively by the model Q_{lm} and S_{II} is the expectation of the power spectrum of I_{lm}. Thus, gravitational measurement noise will bias the gravitational power spectrum by an additive constant.

In quantifying the relationship between gravity and topography, it is useful to work with ratios of their cross power spectra. One quantity, referred to as the *admittance*, is defined by the cross power of the two functions divided by the power of the topography (e.g., Dorman and Lewis, 1970):

$$Z(l) = \frac{S_{hg}(l)}{S_{hh}(l)} \quad [28]$$

This function is not biased by the presence of noise, and when the transfer function Q_{lm} is isotropic (i.e., independent of m), it is an unbiased estimate of Q_l. Another ratio that is commonly used and that is dimensionless is the *correlation*:

$$\gamma(l) = \frac{S_{hg}(l)}{\sqrt{S_{hh}(l) S_{gg}(l)}} \quad [29]$$

If the coefficients h_{lm} and g_{lm} possess a zero mean, the aforementioned function is simply the correlation coefficient of the two sets of harmonic coefficients for a given degree l, which is bounded between 1 and -1. Whereas noise will not bias the admittance function, it will bias the degree correlation function toward lower values as a consequence of the gravitational power spectrum in the denominator of this expression. The term *coherence* is usually reserved for the correlation squared, but this definition is not universally followed. Since squaring the correlation discards information (its sign), its use is not advocated here. It is noted that the isotropic version of eqn [26] predicts that the spectral correlation coefficient should always be ± 1. Nevertheless, as discussed in Section 10.05.7, nonisotropic models yield expressions that are degree-dependent. Whereas other ratios of cross power spectra could be constructed, only two will be independent. Since the relationship between gravity and topography is nonlinear (see eqn [33]), in the following section, it will be useful to plot the correlation between the observed gravity and the gravity predicted from surface topography.

10.05.3 The Data

10.05.3.1 Earth

10.05.3.1.1 Topography

Despite the fact that the measurement of the Earth's topography and bathymetry has been the subject of numerous government-supported campaigns, large portions of the Earth's surface, namely, the oceans, remain poorly characterized. Indeed, from a global perspective, it can be said that the shapes of Mars and the Moon are known better than that of the planet we call home. Until recently, even the elevations of the aerial portions of the continents possessed long-wavelength uncertainties, a result of mosaicking numerous elevation models, each with its own reference surface, along artificial political boundaries. While major advances have been made in the past decade toward generating global models, the main deficiency is still the sparse bathymetry of the oceans.

Numerous topographic models of the Earth's landmass have been assembled from various sources over the past few decades, including ETOPO1 (1 arc minute resolution; Amante and Eakins, 2009), GLOBE (30 arc second resolution; Hastings and Dunbar, 1999), and GTOPO30 (detailed documentation for these and all following topography models can be found at the appropriate web addresses listed in Table 1). Currently, the most accurate model of the Earth's landmass comes from radar interferometric data collected by the Shuttle Radar Topography Mission (SRTM). During its 10 days of operation onboard the US space shuttle in year 2000, this mission mapped nearly 80% of the landmass between 60°N and 54°S with a horizontal sampling of 1 arc second (~30 m) and an absolute vertical accuracy better than about 10 m (Rabus et al., 2003; Rodríguez et al., 2005). Because of the 5.6 cm wavelength of the radar, elevations correspond to the top of the canopy when vegetation is present.

The bathymetry of the oceans has been measured from ship surveys using echo sounding for over half a century. Unfortunately, the ship tracks sometimes possess large navigational errors, and large data gaps exist. As reviewed by Marks and Smith (2006), many datasets exist that are based upon these measurements, but each has its own peculiarities. In the absence of additional ship survey data, one method that can be used to improve the bathymetry of the oceans is by combining ship survey data with marine gravity data. As is detailed by Smith and Sandwell (1994, 1997), marine gravitational anomalies (as obtained from altimeter-derived sea surface slopes) are highly correlated with seafloor topography over a restricted wavelength band. By combining predicted topography from

bandpassed gravitational anomalies with the long-wavelength bathymetry from shipboard sounding, it has been possible to obtain near-global 2 arc second resolution estimates of seafloor topography. Nevertheless, it should be emphasized that such data products are indeed estimates and that they may be inappropriate for certain types of geophysical modeling.

Global Earth's topography models that combine oceanic bathymetry and landmass topography include ETOPO1 and SRTM30_PLUS. The latter of these represents a combination of the SRTM 30 arc second data and the Smith and Sandwell (1997)-predicted bathymetry, with gaps filled by GTOPO30 data, among others. Using the SRTM, SRTM30_PLUS, and ETOPO1 datasets, a suite of high-degree spherical harmonic models has been constructed by Hirt et al. (2012) (Earth2012) that include models of planetary shape, relief with respect to the geoid, and rock-equivalent topography. An image of this global topographic model is displayed in **Figure 1**, where the elevations are plotted with respect to the EGM96 geoid, which is a good approximation of mean sea level (see Internet documentation in **Table 1** for precise definitions). To obtain absolute radii of Earth (as is necessary for most methods that calculate gravity from topography), it is necessary to add the elevation of the geoid, which is referenced to the WGS84 ellipsoid, and the shape of the WGS84 ellipsoid itself. The WGS84 ellipsoid is a good representation of the Earth's zonal shape and possesses a 21 km rotationally induced difference in elevation between the equator and polar axis. A summary of fundamental planetary constants is given in **Table 2**.

10.05.3.1.2 Gravity

The gravitational field of Earth has been mapped by several techniques, including analyses of satellite radio tracking data, ranging between co-orbiting spacecraft, measurement of gravitational gradients from orbit, and terrestrial measurement campaigns. Satellite altimetry of the ocean surface provides additional constraints, as geoid slopes are proportional to the gravity field in the spectral domain (see, among others, Hwang and Parsons, 1996; Sandwell and Smith, 1997). The construction of global high-resolution models consists of combining the long-wavelength information from satellite tracking data with the short-wavelength information in the terrestrial and oceanic altimeter surveys. The model EGM96 (Lemoine et al., 1998) has been the standard reference for much of the past decade, and this was recently superseded by the model EGM2008 (Pavlis et al., 2012). EGM2008 is complete to degree and order 2160 and makes use of a global 180 degree and order GRACE-based solution, gravitational anomalies over the ocean derived from satellite altimetry, and terrestrial and airborne gravity data. The EIGEN-6c2 gravity model (Förste et al., 2012), developed to degree and order 1949, is largely consistent with EGM2008 and makes use of GRACE intersatellite ranges, GOCE gravity gradients, LAGEOS satellite laser ranging data, ocean geoid, and terrestrial data.

The analysis of data from the ongoing missions (Challenging Mini-Satellite Payload (CHAMP); launched in 2000; Reigber et al., 2004), GRACE (launched in 2002; Tapley et al., 2004), and GOCE (launched in 2009; Floberghagen et al., 2011; Rummel et al., 2011) is providing highly accurate estimates of the Earth's long-wavelength gravity field (corresponding to spherical harmonic degrees less than about 200; Pail et al., 2011) and time variations in gravity resulting from geologic processes. These satellites all utilize continuous satellite to GPS (global positioning system) tracking data for orbit determination. CHAMP and GOCE are single satellites, but the primary payload of GOCE is a three-axis gravity gradiometer. GRACE consists of two satellites on identical orbits, of which the distance between the two is measured to high accuracy by a microwave communication link. The analyses of data from GRACE have been providing monthly global gravity solutions since 2003, allowing for study of hydrogeologic, postseismic deformation, and postglacial rebound phenomena.

The images of the total gravitational and geoid anomalies, as determined from eqn [25] and the second-order approximation of eqns [22] and [24], are shown in **Figure 1** for the model EMG2008. The total gravitational anomaly is here defined as the total gravitational acceleration on the rotating WGS84 ellipsoid after removal of the predicted 'normal gravity' of a homogeneous ellipsoid (e.g., Hofmann-Wellenhof and Moritz, 2006). Though this terminology is in common usage in the geophysics community, geodesists would call this the total gravity disturbance. (In terrestrial geodesy, the gravitational anomaly is defined as the gravity on the geoid minus the normal gravity on the reference ellipsoid, where the two quantities are calculated at different points in space.) The largest gravitational anomalies are seen to be correlated with topography (such as trenches and seamounts), and the geoid height is found to vary by about 200 m.

10.05.3.1.3 Spectral analysis

Spectral and cross-spectral properties of the shape model Earth2012 RET and the EGM2008 potential are plotted in **Figure 2**. For this shape model, the radius of the ocean floor was increased by 1030/2670 times its depth in order to convert the mass of the overlying sea water into 'rock-equivalent topography.' As demonstrated in the left panel, the power spectrum of the geoid is about five orders of magnitude less than that of the topography, which is a reflection of the low amplitudes of the geoid undulations present in **Figure 1**. The error spectrum of the potential model demonstrates that the coefficients are well determined at low degrees, with relative uncertainties gradually increasing to 100% near degree 1800. As a result of the ~260 km altitude of GOCE and the ~400–500 km altitudes of GRACE and CHAMP, the contribution to the gravitational field from these spacecraft is necessarily limited to degrees less than about 250; the higher-degree terms are constrained almost entirely by surface measurements.

The admittance and correlation spectra between the radial gravity and topography are plotted in the right panel of **Figure 2**. Given that the relationship between gravity and topography is nonlinear, especially at high degrees, the correlation spectrum is here calculated using the spherical harmonic coefficients of the free-air gravity and the gravity predicted from topography. The correlation for many of the lowest degrees is seen to be small and in some cases negative. The correlation increases beyond degree 12 from about 0.7 to attain a nearly constant value close to 0.9 near degree 500. The slight decrease in correlation that is observed beyond degree 1800 is likely a result of the sparse gravity coverage over portions of the oceanic and continental crust. If the correlation were calculated using the observed topography,

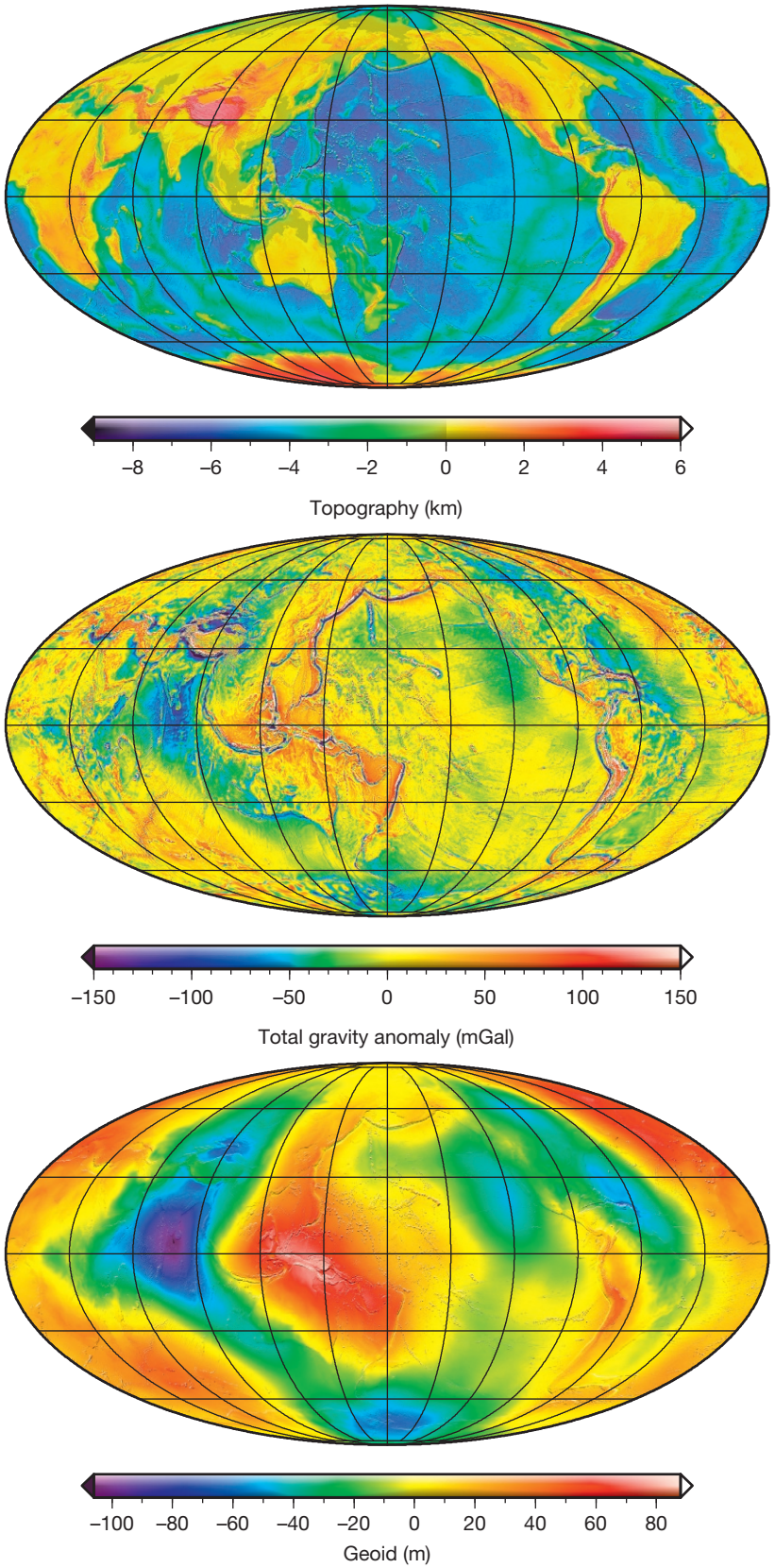


Figure 1 (Top) Global topography and bathymetry of Earth (model Earth2012), referenced to the WGS84 EGM96 geoid, (middle) the EGM2008 total gravitational anomaly on the WGS84 ellipsoid, and (bottom) the height of the EGM2008 geoid above the WGS84 ellipsoid. All images are in a Mollweide projection with a central meridian of 180°W longitude and are overlain by a gradient image derived from the topographic model.

Table 2 Gravitational and shape constants

Parameter	*Value*	*Source*
G	$6.67384(80) \times 10^{-11}$ m^3 kg^{-1} s^{-2}	Committee on Data for Science and Technology; Mohr et al. (2012)
Earth		
GM	$398.6004415 \times 10^{12}$ m^3 s^{-2}	EGM2008; Pavlis et al. (2012)
Semimajor axis	6378.137 km	WGS84; National Imagery and Mapping Agency (2000)
Semiminor axis	6356.7523142 km	WGS84; National Imagery and Mapping Agency (2000)
Radius of sphere of equal volume	6371.00079 km	WGS84; National Imagery and Mapping Agency (2000)
ω	72.921150×10^{-6} rad s^{-1}	WGS84; National Imagery and Mapping Agency (2000)
Venus		
GM	$324.858592 \times 10^{12}$ m^3 s^{-2}	MGNP180U; Konopliv et al. (1999)
Mean planetary radius	6051.878 km	VenusTopo719; this chapter
ω	-299.24×10^{-9} rad s^{-1}	Konopliv et al. (1999)
Mars		
GM	$42.8283748313 \times 10^{12}$ m^3 s^{-2}	JGMRO_110c; Konopliv et al. (2011)
Mean planetary radius	3389.500 km	MarsTopo2600; this chapter
ω	$70.8821807 \times 10^{-6}$ rad s^{-1}	Jacobson (2010)
Mercury		
GM	$22.031839224 \times 10^{12}$ m^3 s^{-2}	GGMES_20v04; Smith et al. (2012)
Mean planetary radius	2439.372 km	GTMES_24v01
ω	$1.2400172589 \times 10^{-6}$ rad s^{-1}	Rambaux and Bois (2004)
The Moon		
GM	$4.90280007 \times 10^{12}$ m^3 s^{-2}	DE430; Williams et al. (2014)
Mean planetary radius	1737.151 km	LOLA2600p; this chapter
ω	2.6617073×10^{-6} rad s^{-1}	Yoder (1995)

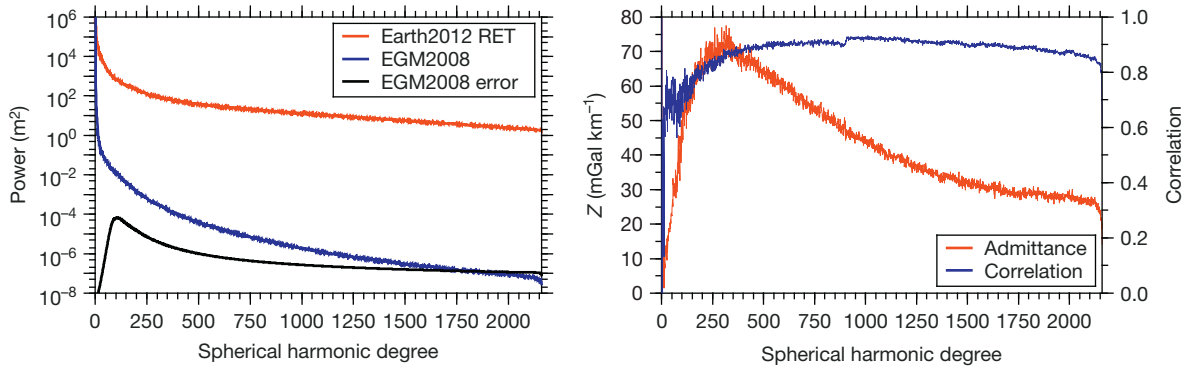


Figure 2 Power and cross power spectra of the Earth's radial gravity and topography. (Left) Power spectra of the topography (Earth2012 RET), geoid (EGM2008), and geoid error. (Right) Admittance and correlation spectra of the radial gravity and topography. The correlation spectrum is calculated using the spherical harmonic coefficients of the observed free-air gravity and gravity predicted from topography.

instead of gravity predicted from topography, the correlation would decrease to less than 0.5 at the highest degrees.

If the surface topography were completely uncompensated, which should be a good approximation beyond degree 300, the admittance would have a near-constant value of $2\pi\rho G$, or 42 mGal km^{-1} times the surface density in units of g cm^{-3}. The observed function is everywhere lower than this theoretical value by about a factor of 2. The low admittances at high degrees are a consequence of the fact that the gravity and surface topography are not perfectly correlated on a global scale as a result of subsurface loads (see discussion in **Section 10.05.6.2**).

10.05.3.2 Venus

10.05.3.2.1 Topography

The planet Venus possesses a dense atmosphere and is perpetually enshrouded by opaque clouds of sulfuric acid. In order to obtain measurements of the surface, it is necessary to make use of electromagnetic frequencies, such as microwaves, where the atmosphere is transparent. Surface elevations of Venus have been measured from orbit using radar altimeters onboard the missions Pioneer Venus Orbiter, Venera 15 and 16, and Magellan. Of these, the Magellan spacecraft, which was in orbit between 1990 and 1994, collected the highest resolution measurements on a near-global scale (for a detailed description, see Ford and Pettengill, 1992).

As a result of the elliptical orbit of the Magellan spacecraft, the spatial resolution of the elevation measurements varied between 8×10 km at periapse and 19×30 km at the north pole (Rappaport et al., 1999). Over 4 million range measurements were ultimately collected, and these were used to construct a 5×5 km gridded elevation model. With the exception of a few relatively minor data gaps covering about 2.6% of the planet's surface area, coverage of the planet is fairly uniform.

(If data from previous missions are used to fill the gaps, then just less than 1% of the planet is not covered at this resolution.) Though the spacecraft-surface range measurements are estimated to have an accuracy of less than 10 m (errors can be larger over rough and steeply sloping terrain), uncertainties in the spacecraft orbit at the time of initial processing were sometimes much greater, especially during superior conjunction. The most recent gravity model of Konopliv et al. (1999) has improved considerably the spacecraft navigational errors, and these improved orbit predictions have been used by Rappaport et al. (1999) in a complete reprocessing of the altimetry dataset (archived as GTDR3.2). Horizontal uncertainties in the footprint locations are insignificant in comparison with the footprint size, and the rms radial uncertainty is estimated to be less than 20 m.

An image of the Venusian topography (derived from the spherical harmonic model VenusTopo719; **Section 10.05.3.2.3**) is shown in **Figure 3**, where it is referenced to the geoid. Whereas the hypsometry of Venus is unimodal (e.g., Ford and Pettengill, 1992; Rosenblatt et al., 1994), in contrast to that of Earth that is bimodal, Venus can be broadly characterized by its low-lying plains, 'continental' plateaus, and volcanic swells. The most prominent highlands include Aphrodite Terra, which lies along the equator, and Ishtar Terra, which is located at high northern latitudes. Ishtar Terra and Aphrodite Terra differ in that the former is flanked by high-elevation mountain ranges. Isolated domical volcanic provinces that possess prominent rift valleys include Atla (0°N, 200°E) and Beta (25°N, 280°E) Regiones. The highest topographic excursion corresponds to Maxwell Montes, located in Ishtar Terra, which reaches more than 10 km above the surrounding plains.

10.05.3.2.2 Gravity

Models of the gravitational field of Venus have been constructed through the analyses of tracking data from the Pioneer Venus Orbiter and Magellan spacecraft (for a review, see Sjogren et al., 1997). The orbit of the Pioneer Venus spacecraft was highly eccentric and possessed periapse altitudes as low as 150 km near the equator. The Magellan spacecraft was initially on an eccentric orbit as well, but through the technique of aerobraking during the gravity mapping phase of the mission, the orbit was transformed to a more circular one, with periapse and apoapse altitudes varying between 155–220 and 350–600 km, respectively.

The most recent model of the Venusian gravitational field is the 180 degree and order JPL (Jet Propulsion Laboratory) model MGNP180U of Konopliv et al. (1999). Because of computational constraints at the time, this model was constructed in three phases. In the first step, a model to spherical harmonic degree 120 was generated using the full unconstrained covariance matrix and a spatial a priori constraint that depended on the strength of the gravitational accelerations (such models are labeled SAAP for Surface Acceleration A Priori). The second step used this model as the nominal solution and then solved for the coefficients from degrees 116 to 155 using the same spatial constraint. For the third step, the coefficients were determined from degrees 154 to 180, but instead of using a spatial constraint, the spherical harmonic coefficients were biased toward a global power law (i.e., a 'Kaula rule'). Future models could be improved by performing the inversion in a single step. As a result of the spatial constraint that was employed in the first two steps, the spatial resolution of the model varies dramatically with position on the surface. Spectral resolutions approaching spherical harmonic degree 180 may be realized close to the equator, but other regions possess resolutions as low as degree 40 (see **Figure 3** of Konopliv et al., 1999).

Images of the MGNP180U radial gravitational field and geoid are presented in **Figure 3**, evaluated at a radius of 6051 km, where the spectral coefficients have been truncated beyond degree 65. As a result of the slow retrograde rotation of Venus, there is no appreciable rotational flattening of the planet. These plots show that most gravitational and geoid anomalies are highly correlated with the surface topography. The largest radial gravitational anomalies are associated with the volcanoes Maat and Ossa Mons in Atla Regio, with values reaching about 270 mGal. The high elevations of Maxwell Montes, Beta Regio, and numerous smaller volcanic provinces are also seen to possess significant anomalies. Uncertainties in the radial component of the gravitational field are typically 10 mGal at the surface but can be as high as 50 mGal in places.

Like Earth, the geoid undulations of Venus possess a dynamic range of only ~200 m. The largest geoid anomalies correspond to the volcanic swells of Atla and Beta Regiones and the continental regions of Aphrodite Terra and Ishtar Terra. It is also seen that the plains with the lowest elevations possess negative geoid anomalies. Uncertainties in the geoid are typically 1 m but can reach values as high as 4 m.

10.05.3.2.3 Spectral analysis

A 719 degree and order spherical harmonic shape model (VenusTopo719) of Venus was constructed by the author based on the Magellan GTDR3.2 sinusoidally projected data product (see **Table 1**). Missing nodes were filled by data obtained by the Pioneer Venus and Venera 15/16 missions; the remaining gaps were filled by interpolation using the Generic Mapping Tools (GMT) (Wessel and Smith, 1991) command <pr>surface</pr> with a tension parameter of 0.35, and the spherical harmonic expansion was performed using the sampling theorem of Driscoll and Healy (1994). The resulting power spectrum was found to be insensitive to changes in the tension parameter, and the mean planetary radius varied by about 1 m among the various models. A comparison between this spherical harmonic model and the one of Rappaport et al. (1999) shows that the latter suffers from an increasing loss of fidelity with increasing degree (the degree correlation between the two models is ~0.93 at 360°).

Power spectra of the Venusian topography (VenusTopo719) and geoid (MGNP180U) are shown in the left panel of **Figure 4**. These are similar to those of Earth, with the exception that the amplitudes of the degree-1 and degree-2 topographic terms for Venus are relatively smaller. On a log–log plot (data not shown), a change in slope of the topographic power spectrum occurs near degree 100 (Rappaport et al., 1999). This feature might be real, but it is also possible that it is related to interpolating over data gaps before performing the spherical harmonic expansion. The error spectrum of the geoid is seen to be greater than the geoid itself for degrees greater than 65. Though the global values of the potential coefficients are generally unreliable beyond this degree, it should be noted that the spatial resolution of the gravitational field is a strong function of position on the surface. Discontinuities in the error spectrum

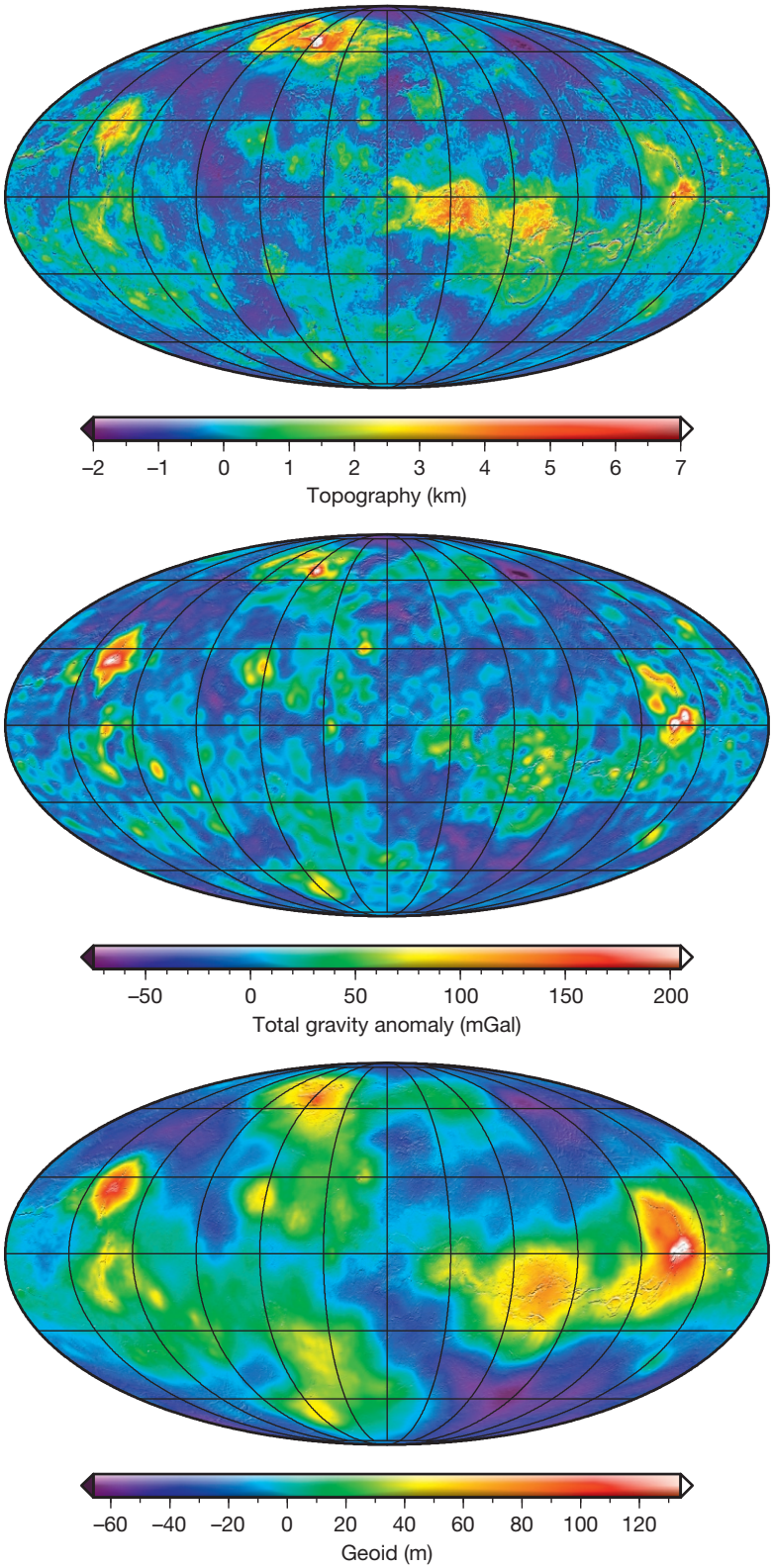


Figure 3 (Top) Global topography of Venus, derived from the spherical harmonic model VenusTopo719, referenced to the geoid. (Middle) Total free-air gravitational anomaly, obtained after truncating the spherical harmonic coefficients of MGNP180U beyond degree 65. (Bottom) Second-order approximation to the geoid. All images are in a Mollweide projection with a central meridian of 60°E longitude and are overlain by a gradient image derived from the topographic model.

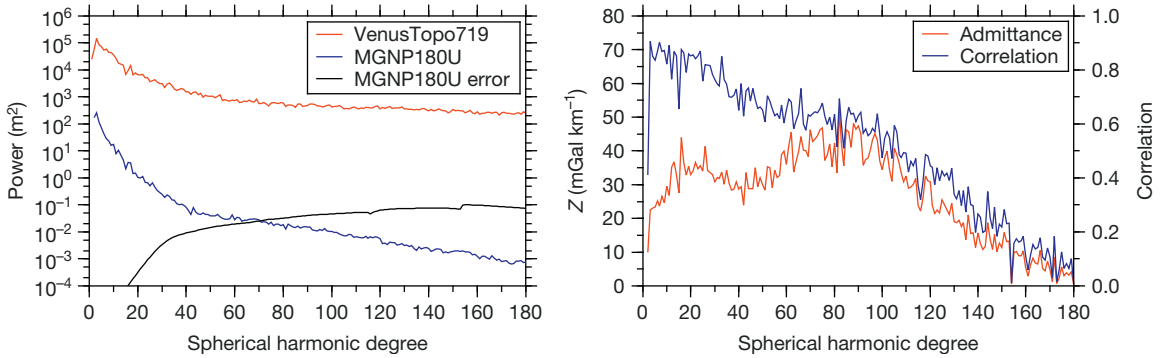


Figure 4 Power and cross power spectra of the gravity and topography of Venus. (Left) Power spectra of the topography (VenusTopo719), geoid (MGNP180U), and geoid error. (Right) Admittance and correlation spectra of the radial gravity and topography. The correlation spectrum is calculated using the spherical harmonic coefficients of the observed free-air gravity and gravity predicted from topography.

are artifacts of solving for the potential coefficients in three separate steps.

The spectral admittance and correlation functions for the radial component of the gravitational field and topography, plotted in the right panel of Figure 4, are seen to differ significantly from those of Earth. The admittance possesses values between ~30 and 50 mGal km^{-1} for spherical harmonic degrees up to 100, whereas for Earth, the admittance linearly increases from ~0 to 70 mGal km^{-1} up to degree 250. The correlation between the gravitational field and topography is also significantly higher for degrees less than 40 than it is for Earth. Nevertheless, beyond about degree 100, the spectral correlation and admittance are seen to decrease linearly with increasing degree, which is a result of the poor determination of the global potential coefficients. It is of note that both the admittance and correlation for degree 2 are significantly smaller than the neighboring values. As these are unaffected by the slow retrograde rotation of the planet, these low values may demand an origin that is distinct from the higher degrees.

Because the Pioneer Venus Orbiter and Magellan spacecraft were in near-polar orbits, the gravity field is better determined for the near-sectoral terms. Sectoral terms correspond to the case $|m|=l$, for which the corresponding spherical harmonic functions do not possess any latitudinal zero crossings. By considering only those coefficients where $l-|m|<20$, Konopliv et al. (1999) had shown that both the admittance and correlation between gravity and topography are considerably greater than when considering all coefficients combined. In particular, the correlation function remains close to 0.7 for degrees up to 140, at which point it decreases substantially. Thus, while high-degree localized spectral analyses may be justified on Venus, the fidelity of the spectral estimates will be a strong function of both position and the spherical harmonic degree and order.

10.05.3.3 Mars

10.05.3.3.1 Topography

Prior to the 1990s, the best Martian topographic models were constructed by a combination of Earth-based radar data, spacecraft radio occultations, and stereo and photoclinometric observations, all of which suffered from either large uncertainties or a limited spatial extent (for a review, see Esposito et al., 1992). The laser altimeter onboard the MGS spacecraft (MOLA, Mars Orbiter Laser Altimeter) has since collected an impressive dataset that has revolutionized studies of the Martian surface (see Smith et al., 2001a, 1999; Zuber et al., 2000b).

MOLA made more than 640 million ranges to the surface over the period of 4 years, after being inserted into orbit in 1997. The spot size of the laser at the surface was ~168 m, and these were spaced every 300 m in the along-track direction of the spacecraft orbit. The intrinsic range resolution of the instrument was 37.5 cm, but range precision decreases with increasing surface slope and could be as poor as 10 m for slopes near 30°. Whereas the along-track orbit errors are less than the size of the laser footprint, radial orbit errors could sometimes be as high as 10 m. Nevertheless, these orbit-induced errors could be minimized by the use of altimetric crossovers. Crossovers occur whenever two altimeter ground tracks intersect, and the difference in the two elevation measurements is largely a reflection of errors in the orbit determination. By parameterizing these uncertainties by a slowly varying function, the crossover residuals can be minimized (Neumann et al., 2001). Such a procedure was capable of reducing the rms crossover residuals from about 8.3 to 1.8 m. Using these methods, it has been possible to measure temporal variations in CO_2 snow depth that can reach 2 m in the polar regions (Smith et al., 2001b).

The topography of Mars (as determined from the 2600-degree spherical harmonic model MarsTopo2600; see in the succeeding text) referenced to the geoid (calculated to second-order accuracy) is displayed in Figure 5. Two of the most remarkable features are the dichotomy in elevation between the Northern and southern hemispheres and the regionally high elevations of the Tharsis volcanic province that is centered near the equator at 100°W longitude. These two features give rise to a 3.3 km offset of the center of figure from the center of mass that is directed toward 64°S and 99°W, of which the longitudinal offset is directed toward the Tharsis province. In addition to these long-wavelength features, there is also an ~20 km difference between the polar and equatorial radii that is a result of the planet's rotational flattening.

Other major topographic features include the giant Hellas (40°S, 65°W), Argyre (50°S, 40°W), and Isidis (15°N, 85°E)

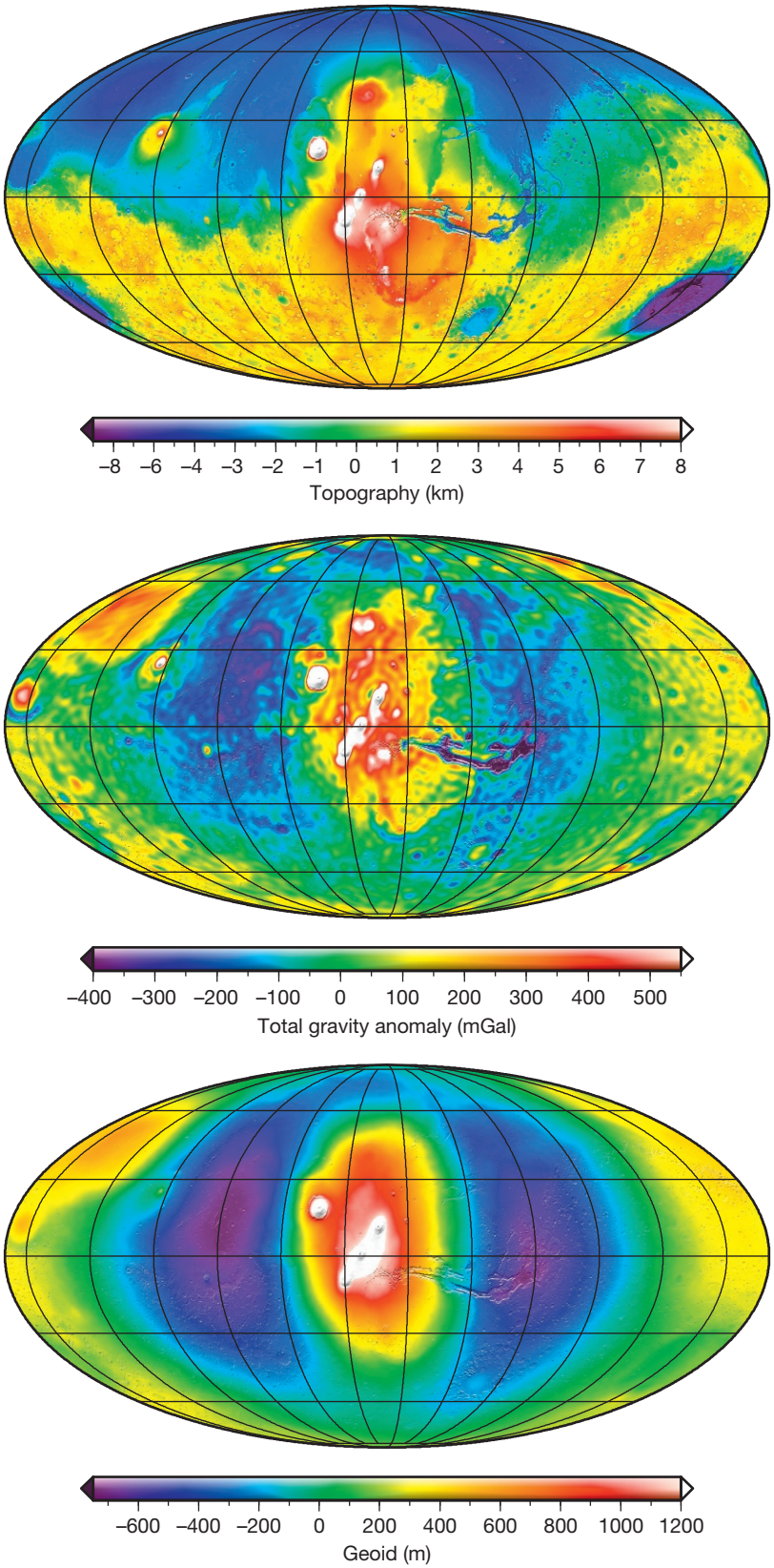


Figure 5 (Top) Global topography of Mars derived from the spherical harmonic shape model MarsTopo2600, referenced to the geoid. (Middle) Total free-air gravitational anomaly on the rotating reference ellipsoid of Ardalan et al. (2010), obtained after truncating the spherical harmonic coefficients of JGMRO_110C beyond degree 85. (Bottom) Second-order approximation to the geoid height with respect to the reference ellipsoid. All images are in a Mollweide projection with a central meridian of 100°W longitude and are overlain by a gradient image derived from the topography model.

impact basins; the Elysium volcanic province (25°N, 145°E); the rift valley Valles Marineris; and the prominent volcanoes that are superposed on the Tharsis province. The highest elevation corresponds to the volcano Olympus Mons, which rises almost 22 km above the MOLA reference geoid.

10.05.3.3.2 Gravity

The gravitational field of Mars has been improved by the successive acquisition of radio tracking data from the Mariner 9, Viking 1 and 2, MGS, Mars Odyssey, and Mars Reconnaissance Orbiter (MRO) missions. A major improvement in the gravity solutions came with the acquisition of data from the MGS mission (see Konopliv et al., 2006; Lemoine et al., 2001; Yuan et al., 2001). This spacecraft was in a near-polar orbit, and during the early portion of the mission when the orbit was highly elliptical, tracking data from altitudes as low as 170 km were acquired at latitudes between 60° and 85°N. Through the technique of aerobraking, the spacecraft was put into a near-circular mapping orbit with periapse and apoapse altitudes of 370 and 435 km, respectively. Mars Odyssey acquired some tracking data from altitudes as low as ~200 km (primarily over the southern hemisphere), with the rest from a near-circular 390 × 455 km mapping orbit. The last major advance was the collection of tracking data from the MRO spacecraft. For this mission, tracking data were acquired from an elliptical orbit with periapse altitudes as low as 255 km over the south pole and apoapse altitudes of 320 km over the north pole. Data were also collected for a brief period with altitudes as low as 220 km over the northern hemisphere.

The most recent and highest resolution gravity solution of Mars is the JPL 110 degree and order model JGMRO_110C (Konopliv et al., 2011). This model employs MGS, Mars Odyssey, and MRO tracking data in combination with surface tracking data from the Pathfinder lander, the latter of which were used to improve knowledge of the orientation of Mars. In the absence of a priori constraints, inversions for the global spherical harmonic coefficients give rise to an unrealistic power spectrum beyond degree 85. In order to obtain a higher-resolution model with realistic power, a Kaula constraint was applied beyond this degree. This solution also obtained the seasonal even zonal J_2 and the better determined odd zonal J_3 signals that result from CO_2 mass exchange between the two polar ice caps and the k_2 Love number from time-variable signals associated with tides raises by the Sun.

An image of the total gravitational anomaly of the model JGMRO_110C is plotted in Figure 5, where the spherical harmonic coefficients have been truncated beyond degree 85. For this map, the normal gravity field predicted for the Mars rotating reference ellipsoid of Ardalan et al. (2010) was removed. Clearly visible are the large positive anomalies associated with the volcanoes in the Tharsis plateau and a negative gravity moat that surrounds this province (Phillips et al., 2001). Large positive anomalies are also evident for some of the largest impact basins, such as Isidis, Argyre, and the buried Utopia basins (45°N, 110°E) that lie in the northern plains. A negative gravitational anomaly is clearly associated with the Valles Marineris rift valley, and negative anomalies surrounding some mountains and volcanoes seem to indicate a flexural origin. The spatial resolution of this field varies laterally, with an effective resolution near spherical harmonic degree 86 for northern latitudes and 110 near the south pole. Formal uncertainties in the radial gravity are at most 10 mGal (A. Konopliv, personal communication).

The Martian geoid, as obtained from the model JGMRO_110C and plotted with respect to the reference ellipsoid of Ardalan et al. (2010), is shown in Figure 5. The geoid undulations of Mars are seen to possess the largest amplitudes among the terrestrial planets, with a dynamic range of over 2.5 km. The signal is largely symmetrical about the Tharsis province, where a central geoid high is surrounded by an annular low. Other geoid highs are associated with the Isidis and Utopia impact basins, as well as the Elysium volcanic rise. Uncertainties in the geoid are generally no more than 2 m (A. Konopliv, personal communication).

10.05.3.3.3 Spectral analysis

The power spectra of the Martian topography (from the author's spherical harmonic model MarsTopo2600, calculated from the gridded datasets available at the PDS (planetary data system) website) and geoid (JGMRO_110C), are plotted in Figure 6. In comparison with the topographic power spectrum, the Martian geoid is seen to have greater amplitudes by about two orders of magnitude than both Earth and Venus. Furthermore, the first 5 degrees of the Martian geoid is considerably greater than would be expected based upon an

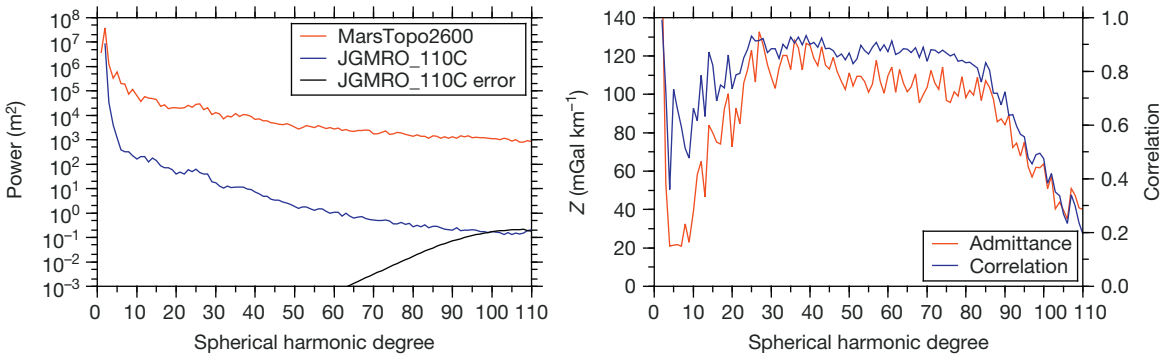


Figure 6 Power and cross power spectra of the gravity and topography of Mars. (Left) Power spectra of the topography (MarsTopo2600), geoid (JGMRO_110C), and geoid error. (Right) Admittance and correlation spectra of the radial gravity and topography. The correlation spectrum is calculated using the spherical harmonic coefficients of the observed free-air gravity and gravity predicted from topography.

extrapolation of the higher-degree terms. This low-degree signal is likely a consequence of the lithospheric load and deflection associated with the Tharsis province (see Phillips et al., 2001; Zuber and Smith, 1997). The error spectrum of the geoid is seen to be larger than the signal for degrees greater than ~100.

The admittance and correlation of the radial gravitational field and topography are shown in the right panel of Figure 6. The admittance function gradually increases with increasing degree, attaining a relatively constant value beyond degree 30. Beyond degree 85, both the admittance and correlation decrease as a result of the poor resolution of the gravity field. While the shape of the admittance function is somewhat similar to that of Earth, it is important to note that the amplitudes are considerably larger at high degrees (~100 mGal km^{-1} in comparison with ~35 mGal km^{-1}). Indeed, these large values are comparable to what would be expected for uncompensated surface topography. One apparent anomaly with the admittance spectrum is the relatively high value of 53 mGal km^{-1} for the degree-three term.

With few exceptions, the correlation between the gravity and gravity predicted from topography is also seen to be very high, with values between about 0.8 and 0.9. Similar to the Venusian gravity model, the near-sectoral terms of the Martian gravity solutions are relatively better determined because of the near-polar orbits of the MGS and MRO spacecraft. When only these near-sectoral terms are used ($l - |m| < 10$), the correlation between the gravity and topography is considerably larger than that shown in Figure 6 for degrees greater than 85 (Konopliv et al., 2011).

10.05.3.4 Mercury

10.05.3.4.1 Topography

Prior to the MErcury Surface, Space ENvironment, GEochemistry, and Ranging mission, MESSENGER (Solomon et al., 2001), the shape of Mercury was only poorly constrained by Earth-based radar ranging to equatorial latitudes (Anderson et al., 1996a; Clark et al., 1988; Harmon et al., 1986) and by a few local elevations models made from Mariner 10 stereo imagery (André et al., 2005). MESSENGER has since acquired a wealth of topographic data by means of satellite laser altimetry (Smith et al., 2010c; Zuber et al., 2008, 2012), stereophotogrammetry (Preusker et al., 2011), optical limb measurements (Oberst et al., 2011), and spacecraft radio occultations (Perry et al., 2011). The spacecraft continues to acquire data in an extended mission.

MESSENGER was launched in 2004, performed three flybys of Mercury in 2008–09, and was placed into a near-polar orbit in 2011. The orbit of the spacecraft during the mapping phase was highly elliptical, with altitudes varying from about 200 km at latitudes near 60°N to about 15 200 km over the southern hemisphere. Because of this, the laser altimeter could range southward only to 15°S. The laser spot size on the surface varied with altitude, from about 15 to 100 m, and the laser shots were separated by about 400 m along track. The near-polar orbit led to dense altimetric coverage near the north pole, decreasing coverage toward the equator. Supplementing these measurements, ~180 radio occultation point measurements of the planet's shape were made when the spacecraft passed behind the planet, with about half of these being located in the southern hemisphere. The processing of stereo imagery has led to the construction of a global elevation model, and a future challenge is to combine the stereo-derived shape models with the laser altimetry and radio occultation data.

The global shape of Mercury derived from stereo imaging (R. Gaskell, personal communication) is shown in Figure 7. As expected from images of the surface, the planet has been shaped largely by impact processes. Several large impact basins, such as the 1500 km diameter Caloris basin (31°N, 160°E), are visible as are numerous smaller craters. The largest basins are not as topographically distinct as those on Mars and the Moon, suggesting that post-basin modification processes have operated after their formation. One clear example of this is the topographic bulge in the northern portion of the Caloris impact basin that is higher than most of its topographic rim (Zuber et al., 2012). Smooth plains are also present, particular at latitudes poleward of about 60°N, that are likely to represent volcanic flows (Zuber et al., 2012). In terms of the planet's long-wavelength shape, the highest elevations are found approximately near 0° and 180° longitude, which corresponds to the axis of the planet's minimum principal moment of inertia (e.g., Davies and Batson, 1975).

10.05.3.4.2 Gravity

The gravity field of Mercury has been constrained by radio tracking of the Mariner 10 and MESSENGER spacecraft. Tracking data from Mariner 10 were acquired only during two flybys, and these allowed for an estimation of the degree-2 gravity coefficients J_2 and C_{22} (Anderson et al., 1987). Radio tracking of the MESSENGER spacecraft has since given rise to a dramatic improvement in our understanding of Mercury's gravity field, with global solutions being developed to spherical harmonic degree and order 20 (Genova et al., 2013; Smith et al., 2012). Nevertheless, as a result of the highly eccentric orbit of the spacecraft, the global models are poorly constrained over the southern hemisphere. Whereas errors in the surface gravity field are less than 20 mGals at latitudes poleward of 30°N, errors are more than 60 mGals at latitudes southward of about 30°S (Genova et al., 2013). Gravity models with higher spatial resolution over the northern hemisphere should be possible by using localized inversion approaches (cf. Han et al., 2009) or with a spatially varying constraint as was done for Venus (Konopliv et al., 1999).

The total gravitational anomaly of the model GGMES_20v04 is plotted in Figure 7, where several positive gravitational anomalies are visible with values approaching 180 mGal. Some of these anomalies are associated with clear topographic signatures, such as the Caloris impact basin and a domical rise in the northern lowlands (Smith et al., 2012), whereas others appear to have little correlation with topography. The volcanic northern lowlands (north of about 60°N) are associated with a regional negative gravitational anomaly, as are the low elevations associated with the large Rachmaninoff impact basin (28°N, 56°E). Large regions of high-standing topography do not possess any significant gravitational anomaly, indicating that they are likely to be isostatically compensated.

The height of the geoid of Mercury, measured with respect to the mean planetary radius, has a dynamic range of about 400 m. As seen in Figure 7, the planet has prominent signals

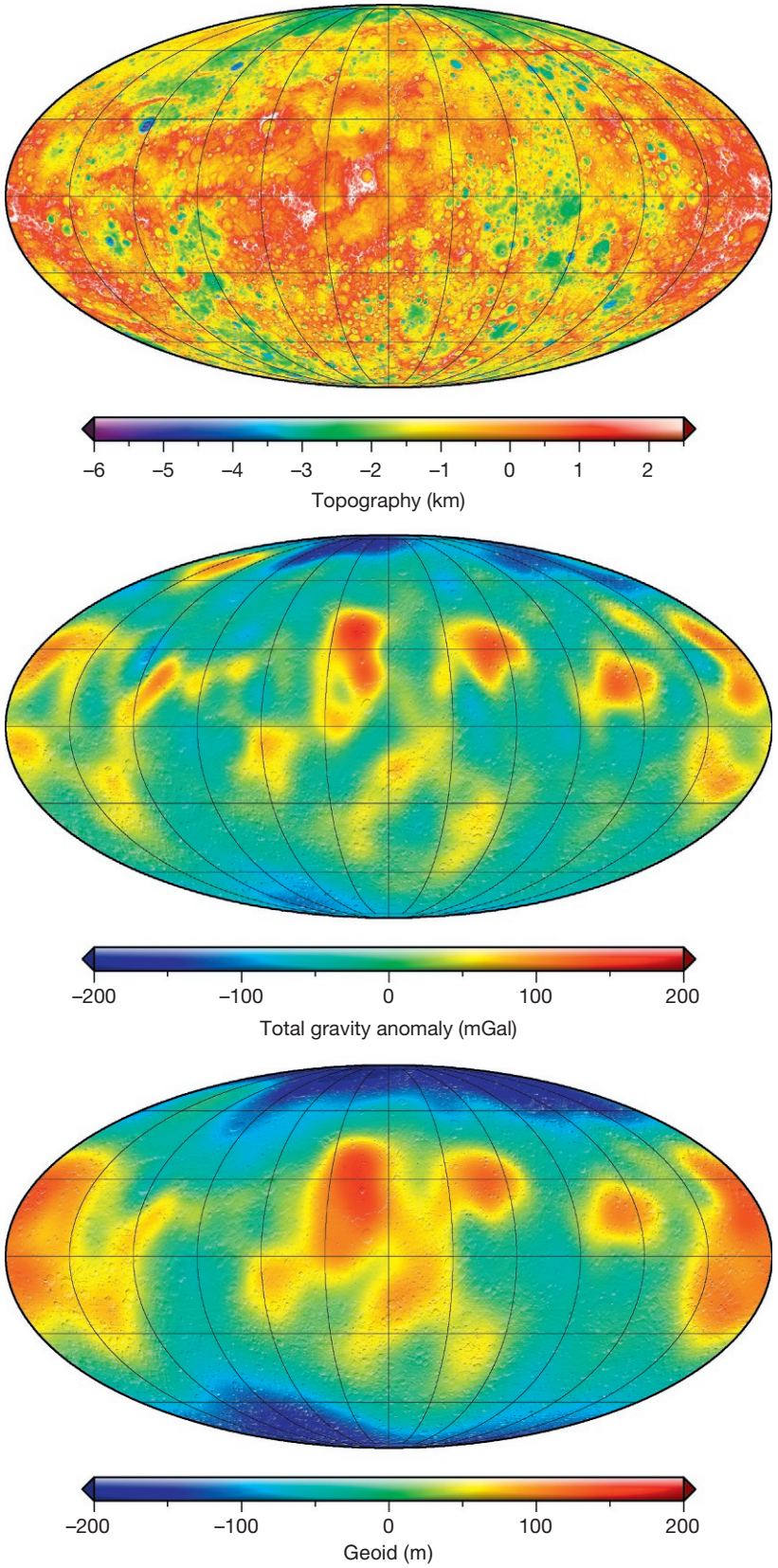


Figure 7 (Top) Global topography of Mercury (R. Gaskell, personal communication), referenced to the geoid. (Middle) Total free-air gravitational anomaly from the spherical harmonic model GGMES_20v04. (Bottom) Second-order approximation to the geoid. All images are in a Mollweide projection with a central meridian of 180°E longitude and are overlain by a gradient image derived from the topographic model.

that arise from the J_2 and C_{22} gravity coefficients. This corresponds to positive geoid anomalies located near 0° and 180° longitude and negative anomalies near both poles, reminiscent of that of the Moon. The C_{22} coefficient is proportional to the difference in the intermediate and minimum principal moments of inertia (e.g., Lambeck, 1988) and is responsible for Mercury having been captured in its present 3/2 spin–orbit resonance (cf. Correia and Laskar, 2009).

10.05.3.4.3 Spectral analysis

A spherical harmonic model of the shape of Mercury was constructed from the MESSENGER laser altimeter and radio occultation data (GTMES_24v01; Smith et al., 2012). The power spectra of this model, the geoid from the model GGMES_20v04, and the geoid error are plotted in **Figure 8**, and the relative amplitudes of the topography and geoid spectra are seen to be comparable to those of Mars. As the error spectrum of the geoid does not intersect that of the geoid, it should be possible to construct higher-resolution models over the northern hemisphere.

The correlation and admittance spectra of the gravitational field and topography are shown in the right panel of **Figure 8**. In this plot, the correlation spectrum was calculated using the observed free-air gravity and the gravity predicted from the surface topography. With the exception of degree 4, the admittance possesses values of a few tens of mGal km^{-1}. This is much less than the value predicted for uncompensated surface topography, indicating that much of the topography at these wavelengths is probably compensated. The correlation possesses values between about 0.4 and 0.8, which is somewhat similar to Mars and the Moon over the same wavelength range. The less than unit correlation is likely a consequence of the presence of mascon basins, such as Caloris, where the gravity and topography are locally anticorrelated, as well as the presence of gravitational anomalies from the mantle and deep crust that do not correlate with surface topography.

10.05.3.5 The Moon

10.05.3.5.1 Topography

The topography of the Moon has been measured by several means, including satellite altimetry, stereophotogrammetry, and radar interferometry (see Wieczorek et al., 2006, for an historical review). However, because of the Moon's synchronous rotation, most early studies were restricted either to the nearside hemisphere or along the equatorial ground tracks of the Apollo command and service modules. The Clementine mission, launched in 1994, was the first to measure absolute elevations of the lunar surface on a near-global scale (Smith et al., 1997; Zuber et al., 1994). 72 548 valid altimetric ranges were acquired along north–south profiles equatorward of ~80°, and in combination with elevation models of the polar regions based on stereophotogrammetry, a global shape model was created (U.S. Geological Survey, 2002). The next major advance came from laser altimetric ranging from three orbiting spacecraft: Kaguya (Araki et al., 2009) and Chang'e 1 (Ping et al., 2009), both launched in 2007, and LRO launched in 2009 (Smith et al., 2010a,b). In addition to including laser altimeters, the acquisition of stereo imagery from these missions has led to the construction of both regional and global terrane models (e.g., Haruyama et al., 2008; Scholten et al., 2012).

The most successful of the recent lunar altimeters is the Lunar Orbiter Laser Altimeter (LOLA) (Smith et al., 2010a,b) onboard the LRO mission. LOLA is a multibeam laser altimeter that operates by spitting a single laser pulse into five output pulses. A five-spot cross pattern is formed on the surface by constructive interference in the far field, which is then detected by five individual receivers, each with a range accuracy of about 10 cm. The spot size of each laser on the surface is 5 m, and the distance between spots is 25 m. The orbital motion of the spacecraft, combined with the 28 Hz repetition frequency, ensures that each five-spot pattern is repeated every 50 m. Bidirectional surface slopes can be calculated for each laser shot or for a variety of longer baselines (Rosenburg et al., 2011). At the present time, more than 6 billion valid ranges have been made to the Moon, and an additional 2 billion are expected over the lifetime of the laser. Since the orbit of the Lunar Reconnaissance Orbiter is polar, the altimetric coverage near the poles is extremely dense. Nevertheless, some kilometer-scale data gaps exist at equatorial latitudes. Stereo imagery can be used to create elevation models with resolutions approaching about 10 m (Haruyama et al., 2008) to 75 m (Scholten et al., 2012), and a future challenge is to combine the altimetric and stereo-based elevation models.

It should be noted that there are two reference frames that are in common use for the Moon (Archinal et al., 2011a,b).

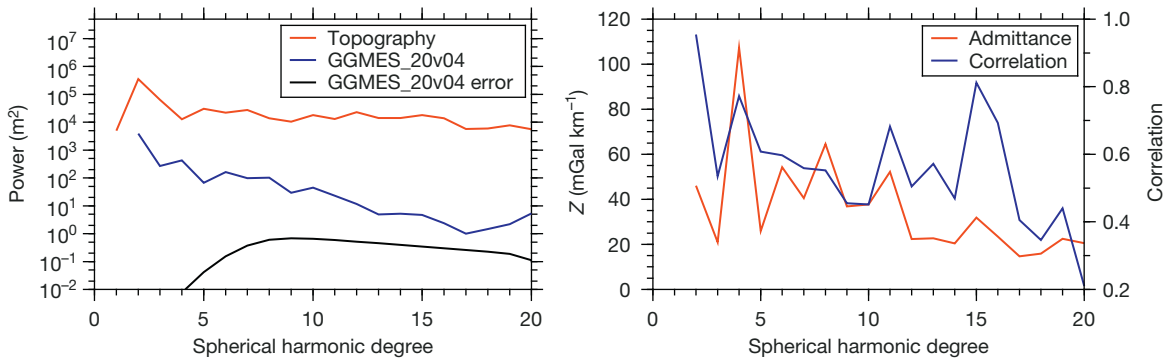


Figure 8 Power and cross power spectra of the gravity and topography of Mercury. (Left) Power spectra of the topography (GTMES_24v01), geoid (GGMES_20v04), and geoid error. (Right) Admittance and correlation spectra of the radial gravity and topography. The correlation spectrum is calculated using the spherical harmonic coefficients of the observed free-air gravity and gravity predicted from topography.

One corresponds to the principal axes defined by the moments of inertia, which is used in the construction of gravitational models. The other is defined by the mean position of Earth and the Moon's rotation axis and is used for many topographic models. The two coordinate frames are mostly offset by a bias with maximum differences up to 873 m, which turns out to be significant when analyzing the high-resolution GRAIL gravity data (Williams et al., 2013). The LOLA topography model (using principal-axis coordinates) is presented in Figure 9 where it is referenced to the GRAIL geoid.

The most dramatic feature of the Moon's topography is the giant South Pole–Aitken impact basin on the farside southern hemisphere. This impact basin possesses a total relief of more than 10 km, and with a diameter of over 2000 km, it is the largest recognized impact structure in the solar system. Other impact basins and craters of various sizes are seen to have shaped the relief of the lunar surface, and the extensive mare basaltic lava flows on the nearside hemisphere, which are relatively younger, are seen to be comparatively smooth. Also of note is that the Moon possesses a 1.9 km displacement of its center of figure from its center of mass in the direction of 8°N and 157°W (e.g., Smith et al., 1997). The total dynamic range in lunar elevation is just under 20 km.

10.05.3.5.2 Gravity

The gravitational field of the Moon has been determined by the radio tracking of orbiting spacecraft and by intersatellite ranging between co-orbiting spacecraft. Radio tracking data from the Lunar Orbiter, Apollo, Clementine, and Lunar Prospector missions led to the development of a degree-150 spherical harmonic model (Konopliv et al., 2001). However, even though the nearside field was moderately well resolved by the 30 km altitude Lunar Prospector data, little was known about the farside hemisphere as it is not possible to track a satellite when it is occulted by the Moon. The first direct tracking data over the farside hemisphere came from the Kaguya mission that was launched in 2007. In addition to the main orbiter (with an altitude of about 100 km), this mission had a relay satellite for performing four-way radio tracking of the main satellite over the farside hemisphere and a subsatellite for very long baseline interferometry (Araki et al., 2009; Goossens et al., 2011; Matsumoto et al., 2010). Data from the Kaguya mission provided a gravitational model that was valid globally to about spherical harmonic degree 70.

The last major advance came with the launch of the GRAIL mission in 2011 (Zuber et al., 2013a,b). This mission was very similar to the terrestrial mission GRACE, in that intersatellite ranges between two co-orbiting spacecraft were used to construct a global gravity model. During the 3 months of the primary mission, the two spacecraft had a mean altitude near 50 km, allowing for the construction of models up to spherical harmonic degree 660 (Konopliv et al., 2013b; Lemoine et al., 2013). During the 3 months extended mission, the altitudes of the two spacecraft were reduced to about 25 km, allowing for the construction of models up to spherical harmonic degree 900. Since the satellite orbits were somewhat elliptical, the minimum elevations over the lunar surface range from only a few kilometers to about 30 km. Though the final models are valid globally to about spherical harmonic degree 750, it is possible to perform localized analyses to higher resolutions where the spacecraft elevations were low.

The total gravitational anomaly and geoid of the GRAIL extended mission model GRGM900B (truncated at degree 550) are plotted in Figure 9. The major features of this map include the large positive gravitational anomalies associated with the nearside impact basins, colloquially referred to as 'mascons' (e.g., Muller and Sjogren, 1968), and the negative gravity moats that surround these basins. The majority of the gravity signal is a result of the surface relief of the Moon, but important signals that are uncorrelated with topography are observed in the lunar mare and elsewhere.

In contrast to Earth, which possesses maximum geoid undulations of ± 100 m, the dynamic range of the lunar geoid is more than 1.1 km. When considering phenomena such as basalt flow directions, it is thus necessary to use elevations that are referenced to the geoid. A long-wavelength J_2 and C_{22} signal is evident in Figure 9, with maxima at the sub- and anti-Earth points, and this signal may represent a fossil tidal deformation that was frozen into the lithosphere early in the orbital evolution of the Moon (e.g., Garrick-Bethell et al., 2006; Jeffreys, 1976; Lambeck and Pullan, 1980).

Finally, it is noted that the orientation of the Moon depends upon its three principal moments of inertia, which are related to the degree-2 gravity coefficients (e.g., Lambeck, 1988). For a synchronously rotating satellite with zero orbital inclination and eccentricity, the minimum energy configuration is achieved when the maximum moment of inertia lies along the rotation axis and when the minimum moment coincides with the Earth–Moon direction (see also Peale, 1969; Ward, 1975). A 180° rotation of the Moon about its rotation axis, as might occur following a large impact event (Wieczorek and Le Feuvre, 2009), would thus be equally stable as its current configuration.

10.05.3.5.3 Spectral analysis

The power spectra of the LOLA principal-axis referenced topography (from the author's model LOLA2600p) and the GRAIL GRGM900B geoid are plotted in the left panel of Figure 10. In comparison with Earth, the power spectrum of the lunar geoid is seen to be about two orders of magnitude more important when compared to the topography spectrum. The lunar geoid and error spectra intersect near degree 750, which is the effective global resolution of the model. Nevertheless, since the GRAIL spacecraft possessed altitudes varying from a few to 30 km during the extended mission, the spatial resolution of the model is in some places better or worse than this spherical harmonic degree. The small upturn in the geoid spectrum near degree 850 is a result of higher-degree aliases that are not resolved by the degree-900 model, indicating that it will be possible to construct even higher-resolution models.

The correlation and admittance spectra for the gravity and topography are plotted in the right pane of Figure 10. In this plot, the correlation spectrum was calculated using the observed free-air gravity and the gravity predicted from the surface topography. The correlation and admittance spectra show low values at the lowest degrees (less than about 0.5 and 100 mGal km^{-1}, respectively). Indeed, these spectra possess negative values for some degrees, indicating that gravity

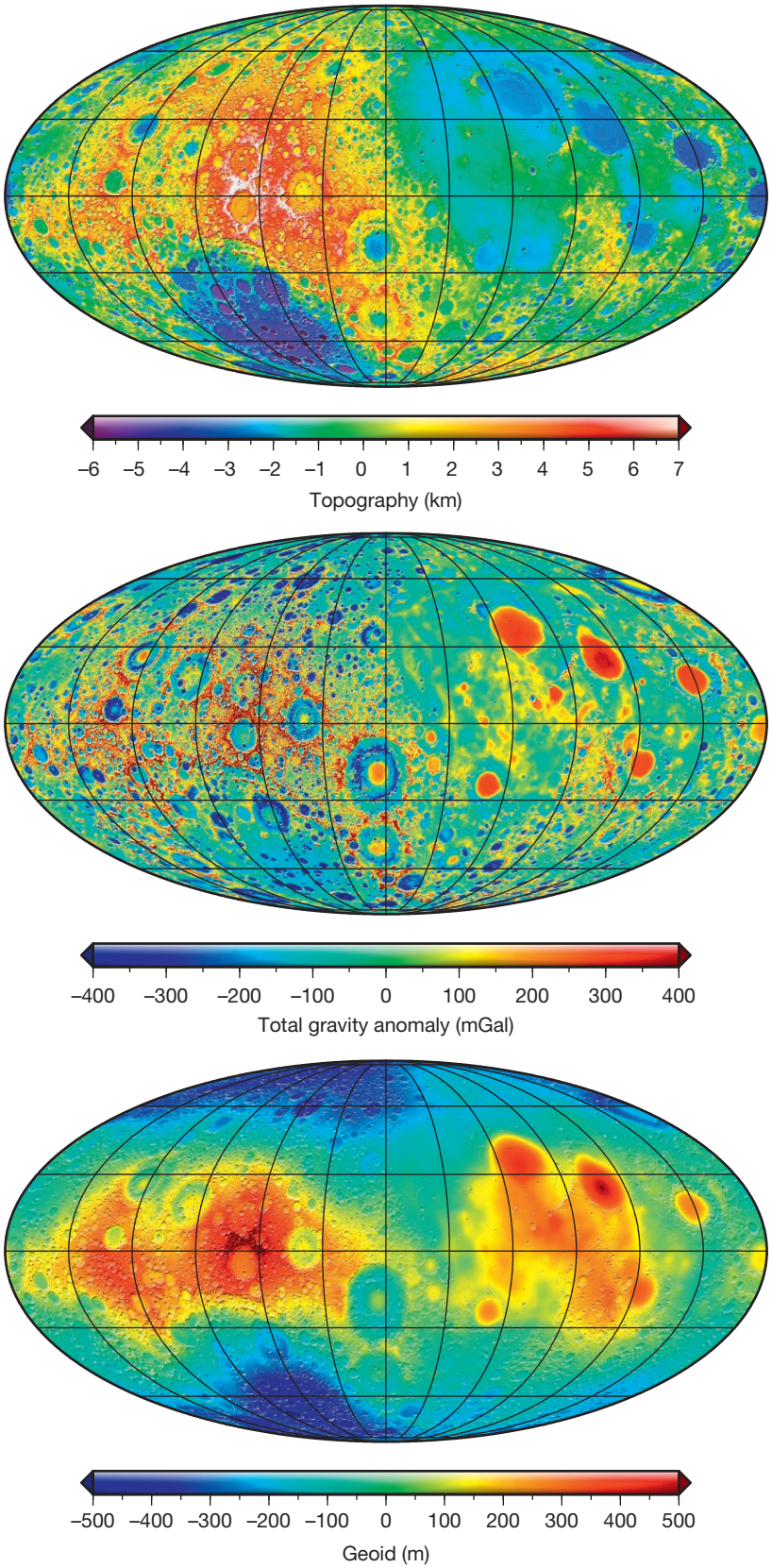


Figure 9 (Top) Global topography of the Moon from the spherical harmonic model LOLA2600p, referenced to the GRGM900B geoid. (Middle) Total gravity obtained from the model GRGM900B evaluated at mean planetary radius after truncating the coefficients beyond degree 550. (Bottom) Second-order approximation to the geoid. All images are in a Mollweide projection with a central meridian of 90°W longitude and are overlain by a gradient image derived from the topographic model. The near- and farside hemispheres are on the right and left halves of these images, respectively.

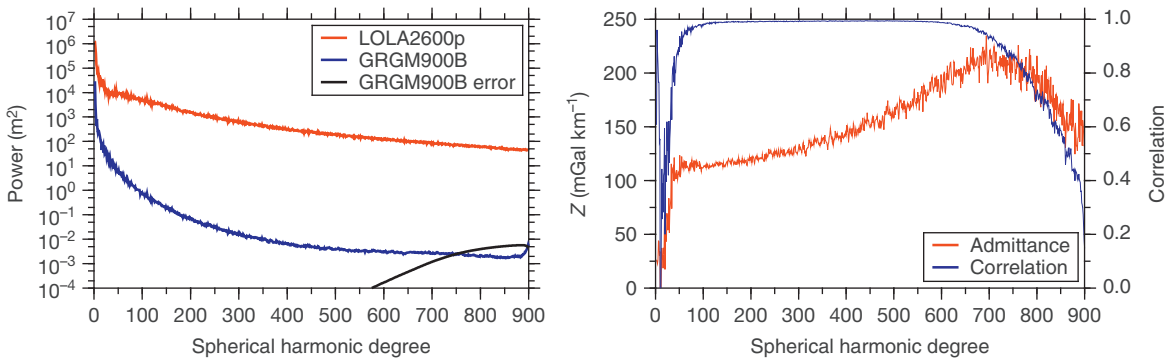


Figure 10 Power and cross power spectra of the gravity and topography of the Moon. (Left) Power spectra of the topography (LOLA2600p), geoid (GRGM900B), and geoid error. (Right) Admittance and correlation spectra of the radial gravity and topography. The correlation spectrum is calculated using the spherical harmonic coefficients of the observed free-air gravity and gravity predicted from topography.

and topography are anticorrelated at these wavelengths. These small values for the lowest degrees are simply a result of the presence of mascon impact basins, which are positive gravitational anomalies that possess low elevations. Beyond about degree 100, the correlation approaches 0.99, and the admittance slowly increases from values near 100 to about 200 mGal km^{-1}. The value of the admittance at these degrees is consistent with uncompensated topography (Wieczorek et al., 2013), and the slow increase in the admittance function is a result of the nonlinear relation between gravity and topography (see **Section 10.05.4**). The correlation starts to decrease near degree 550, and this could be a result of a reduced global fidelity of the gravity model at these degrees or to short-wavelength geologic signals where the gravity and topography signals are not perfectly correlated.

10.05.4 Methods for Calculating Gravity from Topography

Geophysical investigations that employ gravitational and topographic data often attempt to fit the observations with those predicted from a model that contains several parameters. For example, lithospheric flexure calculations depend upon several unknowns, including the effective elastic thickness of the lithosphere, crustal thickness, and the density of the crust and mantle. By comparing predicted gravitational anomalies induced by the deflection of density interfaces with the observed values, the parameters of such a model can be constrained. Whereas several methods exist for calculating the gravitational field related to relief along a density interface, as is described in the succeeding text, this is oftentimes most easily performed in the spherical harmonic domain.

The calculation of the gravitational potential can be considerably simplified by use of the two well-known identities:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|}=\frac{1}{r}\sum_{l=0}^{\infty}\left(\frac{r'}{r}\right)^{l}P_l(\cos\gamma)\quad \text{for}\, r\geq r' \qquad [30]$$

and

$$P_l(\cos\gamma)=\frac{1}{2l+1}\sum_{m=-l}^{l}Y_{lm}(\theta,\phi)Y_{lm}(\theta',\phi') \qquad [31]$$

where P_l is an unnormalized Legendre polynomial and γ is the angle subtended between $\mathbf{r}$ and $\mathbf{r}'$. (Equation [31] is commonly referred to as the Legendre addition theorem.) By inserting these equations into the expression for the gravitational potential (eqn [17]), it is simple to show that the spherical harmonic coefficients of eqn [19] (the *Stokes coefficients*) are equal to

$$C_{lm}=\frac{1}{MR_0^l(2l+1)}\int_V \rho(\mathbf{r}')r'^l Y_{lm}(\theta',\phi')\mathrm{d}V' \qquad [32]$$

It is important to note that this formulation of the gravitational potential can only be used when the observation point is greater than the maximum radius of the body.

Next, consider the case where there is relief $h(\theta, \phi)$ referenced to a spherical interface of radius D and where the density ρ between h and D depends only on latitude and longitude (when h is negative, ρ is considered negative as well). For this situation, it is possible to obtain exact expressions for the corresponding potential coefficients that are similar to those developed by Parker (1972) in the Cartesian domain. Integrating eqn [32] over r' and expanding powers of the relief in a Taylor series, the potential coefficients of eqn [19], referenced to a radius D, are (see Wieczorek and Phillips, 1998)

$$C_{lm}=\frac{4\pi D^3}{M(2l+1)}\sum_{n=1}^{l+3}\frac{(\rho h^n)_{lm}}{D^n n!}\frac{\prod_{j=1}^{n}(l+4-j)}{(l+3)} \qquad [33]$$

The spherical harmonic coefficients of the density multiplied by the relief to the nth power have the explicit expression (cf. eqn [8])

$$(\rho h^n)_{lm}=\frac{1}{4\pi}\int_\Omega[\rho(\theta,\phi)h^n(\theta,\phi)]Y_{lm}(\theta,\phi)\mathrm{d}\Omega \qquad [34]$$

and when the density is constant, eqn [33] reduces to eqn [9] of Wieczorek and Phillips (1998). As a result of the inequality in the identity of eqn [30], this expression for the potential is valid only when the radius r is greater than the maximum elevation of $D+h$. Extensions, special cases, and alternative forms of this equation have been derived independently several times in the literature (e.g., Balmino, 1994; Chambat and Valette, 2005; Chao and Rubincam, 1989; Martinec et al., 1989; Rapp, 1989). It is worth emphasizing that the relief h in eqn [33] must be calculated with respect to a spherical interface and not the ellipsoid or geoid: using a different reference surface would give inaccurate results.

For the case where the density ρ is constant, the potential coefficients can be obtained simply by calculating the spherical harmonic coefficients of the relief to the nth power. While the sum of eqn [33] is finite and hence exact, the number of terms grows linearly with spherical harmonic degree. Nevertheless, as each succeeding term is smaller than the previous, in practice, this sum can be truncated beyond a maximum value n_{max} for which the truncated terms are smaller than the resolution of the gravity model.

For certain applications, it is sufficient to use the first-order term of eqn [33]:

$$C_{lm} = \frac{4\pi D^2 (\rho h)_{lm}}{M(2l+1)} \qquad [35]$$

This expression is commonly referred to as the 'mass-sheet' approximation, as the calculated gravitational anomaly would be exact if it arose from a spherical interface with a surface density of ρh. (The higher-order terms are referred to as the 'finite-amplitude' or 'terrain' correction.) Using this expression, the radial gravity (see eqn [25]) is seen to approach asymptotically with increasing l the Bouguer slab approximation of $2\pi\rho Gh$.

The effect of truncating the sum of eqn [33] beyond n_{max} is illustrated in Figure 11 for the specific case of determining the Bouguer correction of Earth, Venus, Mars, Mercury, and the Moon. The term Bouguer correction here refers to the contribution of the gravitational potential that results from the mass between the mean planetary radius and the surface. The true value of the Bouguer correction was approximated using $n_{max} = 20$, and the maximum difference of the total gravity in the space domain that results from truncating at lower values of n was calculated on the planet's reference ellipsoid. For these calculations, the full resolution of the planetary topography was used in calculating the potential coefficients, but the resulting gravity field was expanded in the space domain only up to the spherical harmonic degree indicated in the figure. To mitigate against aliases that could arise when raising a function to the nth power, all calculations were performed on a grid that could resolve spherical harmonics up to degree 5000.

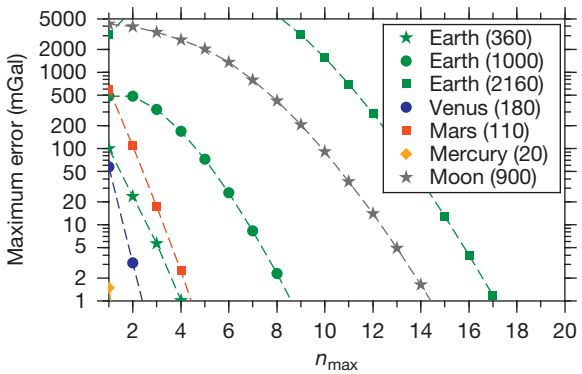


Figure 11 Maximum error associated with the Bouguer correction as a function of the order n_{max}. The Bouguer correction error is here defined as the error in the total gravitational attraction resulting from surface relief with respect to the mean planetary radius, evaluated on the reference ellipsoid. The true value was approximated by $n_{max} = 20$, and the crustal density was assumed to be 2900 kg m^{-3}. The values in parentheses correspond to the maximum spherical harmonic degree used in calculating the gravity field.

Figure 11 shows that the value of n_{max} required to achieve 1 mGal accuracy is highly dependent upon the maximum spherical harmonic degree of the desired gravitational field. To obtain accuracies of a few milligals for Mercury, Venus, and Mars, whose gravity fields are only resolved up to a maximum spherical harmonic degree of 180, only the first four terms of eqn [33] are required. However, when calculating high-resolution gravity fields for the Moon and Earth (with degrees up to 900 and 2160), it is necessary to utilize terms up to 14 and 17, respectively (see also Balmino et al., 2012). If it were necessary to calculate fields up to only spherical harmonic degree 360 for Earth (for use with the gravity model EGM96), then the first four terms would be sufficient.

Finally, it is noted that alternative means exist for calculating the theoretical gravitational field of a body and that these may be preferable to the aforementioned approach for certain applications. One method developed by Belleguic et al. (2005) is quasi-analytic and allows for the calculation of the potential and gravity at any point in a body (this is in contrast to the aforementioned approach that is applicable only to radii greater than the maximum radius). This method starts by mapping irregularly shaped density interfaces to spherical ones and then determines the radial derivatives of the potential and gravitational field on this surface. Using exact values for the potential and gravity field on an interface exterior to the planet (as obtained from a method similar to eqn [33]), these fields are then propagated downward using a first-order Taylor series approximation. This technique is useful for lithospheric flexure calculations as the net lithospheric load is a function of the potential at the major density interfaces.

A second method for calculating the gravitational field is based upon approximating the shape of a celestial body by a polyhedron. Exact expressions for the potential of a homogeneous polyhedron have been derived by Werner and Scheeres (1997), and expressions for the corresponding spherical harmonic coefficients are given in Werner (1997) (see also Tsoulis et al., 2009). The benefit of using this approach is that the resolution of the model (i.e., the spacing between vertices) can be varied according to the resolution of the gravitational field. An application of this method for determining the interior density of an asteroid is given by Scheeres et al. (2000). Alternatively, one could transform eqn [17] into a surface integral by use of Gauss' law and perform the integral numerically for a given shape model (e.g., Cheng et al., 2002).

10.05.5 Crustal Thickness Modeling

It is well known that the modeling of potential fields is nonunique. As an example, eqns [19] and [35] show that any external gravitational field can be interpreted as a surface density ρh placed at an arbitrary radius D. Nevertheless, by using simplifying assumptions based on geologic expectations, it becomes possible to invert uniquely for parameters related to the interior structure of a planet.

Perhaps the simplest example of such an investigation is the construction of a global crustal thickness model. In this case, the nonuniqueness associated with potential modeling is removed by assuming that the observed gravitational field arises only from relief along the surface and crust–mantle

interface and that the densities of the crust and mantle are known. It is further required either to assume a mean crustal thickness or to anchor the inverted crustal thickness to a given value at a specific locale. If lateral density variations in either layer could be constrained by other means, these could be incorporated into the model.

The first step is to calculate the Bouguer correction, which is the contribution to the potential of surface relief referenced to the mean planetary radius. Subtracting this from the observed gravitational field yields the Bouguer anomaly, and this is then ascribed to being caused by relief along the crust–mantle interface at an average radius D. To the first order, this relief could be determined in the spectral domain by downward continuing the Bouguer anomaly coefficients C_{lm}^{BA} to radius D and then setting these equal to those predicted from the mass-sheet approximation of eqn [35]. However, two additional factors generally need to be taken into account in such an analysis. First, downward continuing the Bouguer anomaly amplifies short-wavelength noise that is often present in the observed gravitational field. Second, the first-order mass-sheet approximation may not be sufficiently accurate if the undulations along the crust–mantle interface are large.

By minimizing the difference in the spectral domain between the observed and predicted Bouguer anomalies, the shape of the crust–mantle interface can be computed via the following equation (see Wieczorek and Phillips, 1998):

$$h_{lm} = w_l \left[\frac{C_{lm}^{BA} M (2l+1)}{4\pi \Delta \rho D^2} \left(\frac{R}{D} \right)^l - D \sum_{n=2}^{l+3} \frac{(h^n)_{lm}}{D^n n!} \frac{\prod_{j=1}^{n} (l+4-j)}{(l+3)} \right] \quad [36]$$

The density jump across the crust and mantle is given by $\Delta\rho$, R is the reference radius of the Bouguer anomaly coefficients, and w_l is a filter that stabilizes the downward continuation procedure. The explicit form of the filter can be determined by using an additional constraint that minimizes the mantle relief.

While there is no simple analytic solution to eqn [36], the relief h can be determined using an iterative approach: First, the coefficients h_{lm} are approximated by ignoring the higher-order terms on the right-hand side; then, using this estimate, the higher-order terms are calculated, and a new estimate of h_{lm} is obtained (see also Wieczorek et al., 2013). The second step is iterated until the relief converges to a stable solution. It is important to note that these models do not assume the crust to be isostatically compensated; such a hypothesis could be tested for a given model. Example crustal thickness models that were obtained using this procedure are discussed in the succeeding text, along with the major modeling assumptions that are specific to each body.

Crustal thickness models for Venus and Mars are presented in Figure 12. For the Venusian model, a mean crustal thickness of 35 km was assumed, the potential and topography coefficients were truncated beyond degree 60, and densities of 2900 and 3330 kg m^{-3} were used for the crust and mantle, respectively. The inclusion of finite-amplitude corrections for Venus affects the obtained crustal thicknesses by only a few kilometers. (For alternative models, see Orth and Solomatov, 2012; James et al., 2013.) The model for Mars is an updated version from Neumann et al. (2004) that uses the JGM95J01 gravity model. An average crustal density of 2900 kg m^{-3} was assumed for this model, and the low density of the polar caps, the higher than average densities of the Tharsis volcanoes, and the gravitational attraction resulting from the core flattening were explicitly taken into account. A mean crustal thickness of 45 km was assumed, and in downward continuing the Bouguer anomaly, a filter was constructed such that the power spectrum of the Moho relief resembled that of the surface relief.

A crustal thickness model for Mercury is presented in Figure 13. For this model, densities of 2900 and 3200 kg m^{-3} were used for the crust and mantle, respectively, and the filter w_l was chosen to possess a value of 0.5 at degree 40. Under the assumption that the minimum crustal thickness on Mercury is 5 km, the average crustal thickness is found to be 23 km. Since the spatial resolution of the gravity field is poor in the southern hemisphere, the inverted crustal structure there should be treated with caution. The crustal thickness model for the Moon in this figure is an updated version from Wieczorek et al. (2013) that uses gravity data derived from the GRAIL extended mission. Lateral variations in crustal density, which were estimated from remote sensing data, were used in calculating the gravity field for this model. A crustal porosity of 12% was assumed, a mantle density of 3220 kg m^{-3} was used, and the filter w_l was chosen to have a value of 0.5 at 80°. With an average crustal thickness of 34 km, this model possesses minimum crustal thicknesses that are less than 1 km and fits the 30 km seismic constraint (Lognonné et al., 2003) at the Apollo 12 and 14 landing sites. Neglecting the finite-amplitude terms in eqn [36] would give rise to errors as large as 20 km for the Moon (Neumann et al., 1996).

10.05.6 Admittance Modeling

In the crustal thickness modeling of the preceding section, the nonuniqueness associated with potential modeling was removed by making certain assumptions about the mean crustal thickness and the density of the crust and mantle. These and other parameters can be estimated if one instead assumes that surface topography is supported by a specific mechanism, such as lithospheric flexure or Airy compensation. Using such a model, the relationship between gravity and topography can be determined, and by comparing to the observed values, model parameters can be estimated. As is described in the following two subsections, two methods are in common use: One is based upon modeling the GTR in the space domain, whereas the other models the admittance and correlation functions in the spectral domain.

10.05.6.1 Spatial Domain

One method that has proven to be fruitful for estimating the mean crustal thickness of a planet is modeling of the GTR in the space domain. This technique was initially developed by Ockendon and Turcotte (1977) and Haxby and Turcotte (1978) for Earth, where it was shown that the isostatic geoid anomaly was approximately equal to the vertical dipole moment of density variations within the lithosphere. For the specific cases of Airy and Pratt isostasy, the ratio between the geoid and topography was found to be proportional to

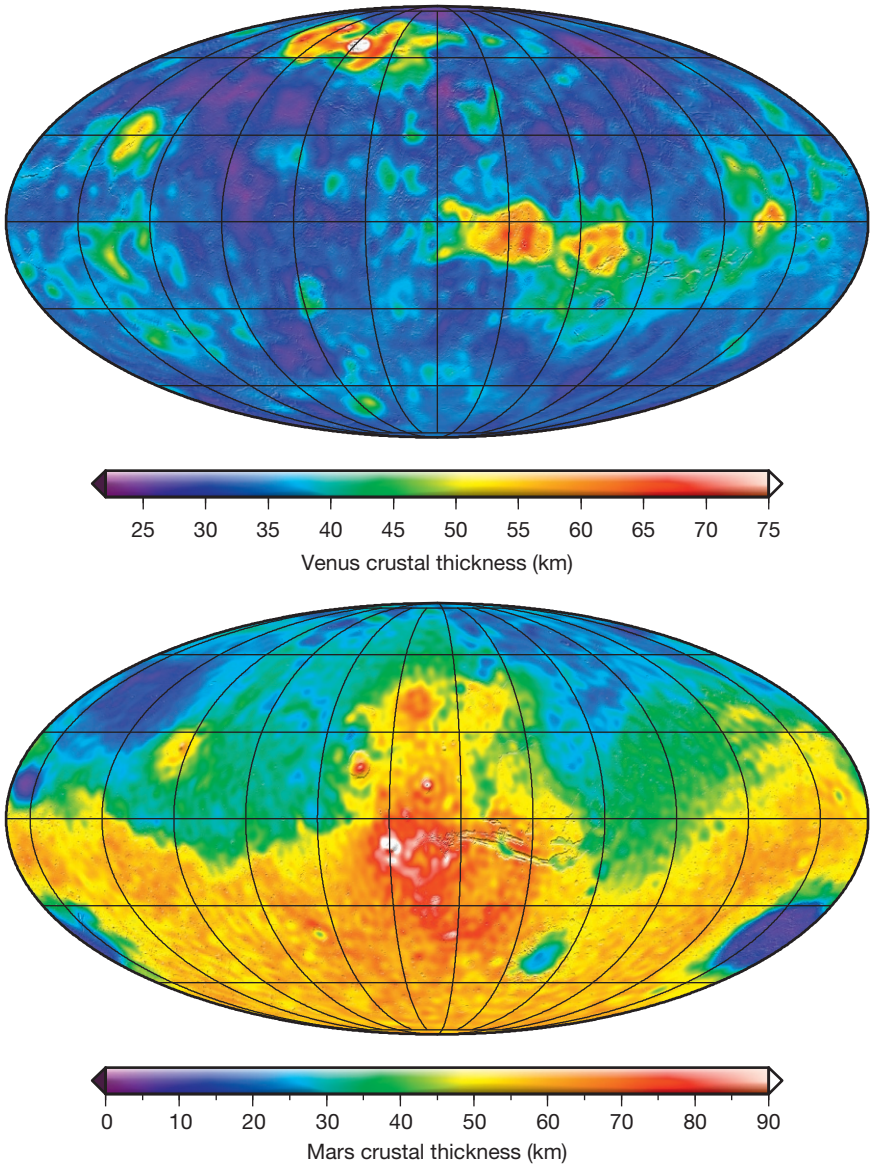


Figure 12 Crustal thickness models for Venus and Mars. Both images are in a Mollweide projection and are overlain by a gradient image derived from the surface topography. The central meridian is 60°E for Venus and 100°W for Mars.

the crustal thickness. Their method was derived using a Cartesian geometry and is strictly valid in the long-wavelength limit.

An alternative approach has been developed in spherical coordinates by Wieczorek and Phillips (1997) where it has been shown that the GTR, N/h, can be approximated by the expression

$$\frac{N}{h} = R \sum_{l=l_{\min}}^{l_{\max}} W_l Q_l \qquad [37]$$

where $l_{\min}$ and $l_{\max}$ correspond to the minimum and maximum spherical harmonic degrees that are considered, R is the mean planetary radius, Q_l is the linear transfer function between the potential and topographic coefficients, and W_l is a weighting function that is proportional to the topographic power of degree l:

$$W_l = S_{hh}(l) \Big/ \sum_{i=l_{\min}}^{l_{\max}} S_{hh}(i) \qquad [38]$$

The underlying assumption of this model is that the GTR is independent of position for a given compensation model, and this has been validated empirically for the highlands of the Moon and Mars for the case of Airy isostasy (Wieczorek and Phillips, 1997; Wieczorek and Zuber, 2004).

As the power spectra of planetary topography are 'red' (i.e., they possess more power at long wavelengths than short wavelengths), eqn [38] shows that the largest contribution to the GTR will come inevitably from the lowest degrees. As an example, approximately 80% of the GTR for the Moon is determined by degrees less than 30. Since the topography of the ancient highland crust of a planet is likely to be isostatically

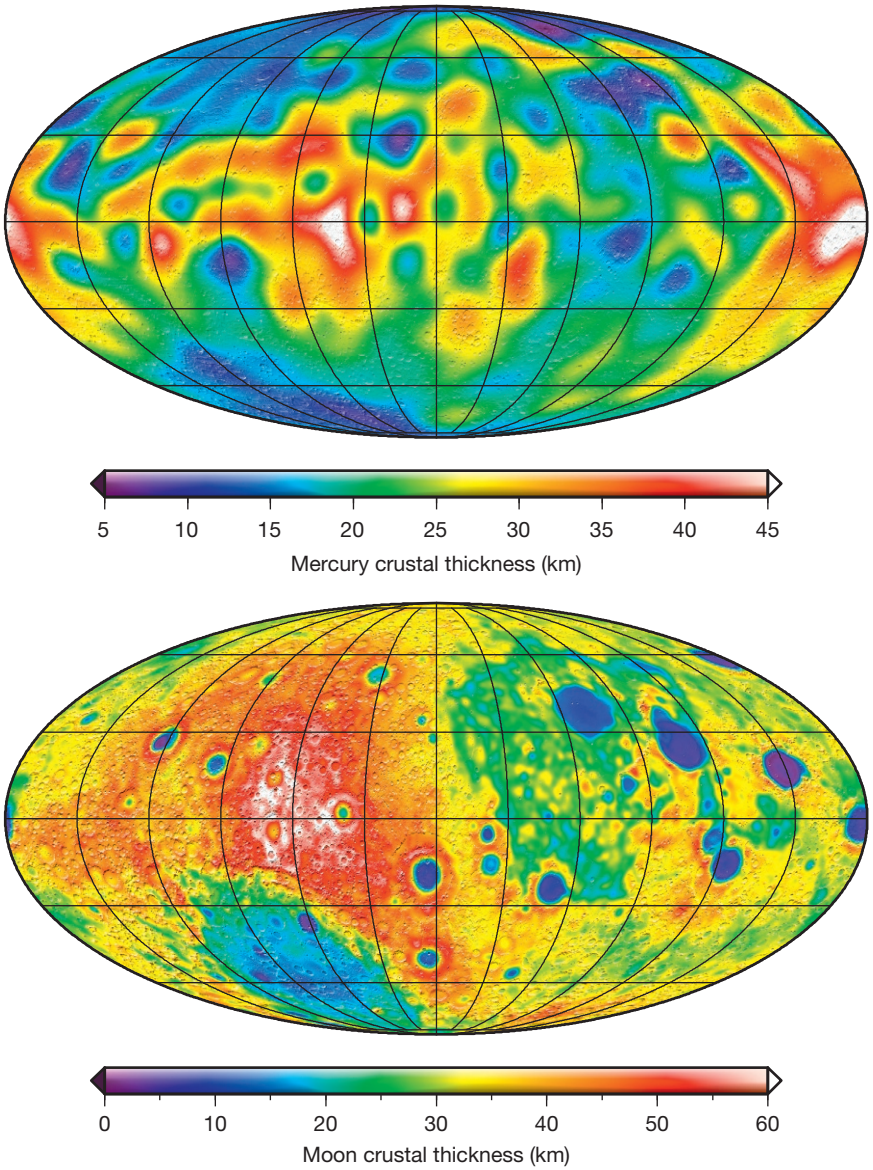


Figure 13 Crustal thickness models for Mercury and the Moon. Both images are in a Mollweide projection and are overlain by a gradient image derived from the surface topography. The central meridian is 180°E for Mercury and 90°W for the Moon.

compensated at these wavelengths, it is common to employ a model based on the condition of Airy isostasy for these regions. Assuming that the density of the crust is constant and using the condition of equal mass in crust–mantle columns, it is straightforward to show using eqn [35] that the transfer function between the potential and topographic coefficients is

$$Q_l = \frac{C_{lm}}{h_{lm}} = \frac{4\pi\rho_c R^2}{M(2l+1)}\left[1 - \left(\frac{R-T_c}{R}\right)^l\right] \qquad [39]$$

where ρ_c is the density of the crust and T_c is its mean thickness.

In practice, the GTR is determined by fitting the observations to a straight line within a region that is believed to be consistent with the employed model. By utilizing a plot of the predicted GTR versus T_c (obtained from eqns [37] to [39]), the crustal thickness can then be estimated for a given value of ρ_c. Nevertheless, as the GTR is influenced strongly by the longest wavelength components of the gravity and topography, several aspects need to be considered carefully when performing such an analysis.

It is first necessary to ensure that the entire signal of the geoid and topography are governed by the same compensation model. This can never be entirely satisfied, but certain anomalous long-wavelength features can sometimes be identified and removed. For instance, most planets and satellites possess significant rotational and/or tidal contributions to their degree-2 shape, and these signatures can be minimized by setting these coefficients to zero. For Mars, in addition to the degree-2 rotational signature, the longest wavelength components have been affected by the load and flexural response associated with the Tharsis province (see Phillips et al., 2001; Wieczorek and Zuber, 2004; Zuber and Smith, 1997).

Furthermore, as the degree-1 potential terms are zero when the gravitational field is referenced to the body's center of mass, any degree-1 topography that exists may need to be treated separately. Finally, as the GTR is determined largely by the longest wavelength components of the geoid and topography, it is necessary that the region of interest be sufficiently large when regressing the geoid and topography data.

10.05.6.2 Spectral Domain

Two major shortcomings associated with modeling the GTR are that only a single wavelength-independent parameter is used and the observed value could be biased by long-wavelength features that are unrelated to the assumed compensation model (such as lateral density anomalies in the mantle). An alternative modeling approach that largely bypasses these concerns is to model the relationship between the gravity and topography for a certain region in the spectral domain. As wavelength-dependent admittance and correlation functions are obtained, it is in principle possible to invert for several model parameters. The major shortcoming of this approach is that low spectral resolutions of planetary datasets can be particularly limiting when performing localized spectral analyses, as described in the following section. In this section, the basic concepts involved with global spectral analyses are described.

Let us presume that the potential and topography coefficients are related via an equation of the form

$$C_{lm} = Q_{lm} h_{lm} + I_{lm} \qquad [40]$$

where Q_{lm} is a linear nonisotropic transfer function and I_{lm} is that portion of the potential not described by the model. For simplicity, it will be assumed that I_{lm} is zero (though this assumption will be relaxed later) and that the topography is noise-free. Even though the aforementioned relationship is inherently nonisotropic, it is often useful to work with the power and cross power spectra of the gravitational field and topography (S_{hh}, S_{gg}, and S_{hg}), which only depend upon spherical harmonic degree l. The goal is to fit these three functions to those obtained from an appropriate model. In order to remove the model dependence of certain nuisance parameters, it is convenient to work with ratios of these three cross power spectra. Although several such ratios involving powers of these are possible, only two will be independent, and it is traditional to use the admittance and correlation spectra as defined previously by eqns [28] and [46]. If a model describing a planet's gravity and topography is to be considered successful, then it must satisfy both of these functions. If one or both of these functions cannot be fit for a given degree, then this is a clear indication that either the model assumptions are inappropriate for the region being investigated or the input datasets are not sufficiently accurate.

If one treats the lithosphere of a planet as a thin elastic spherical shell overlying a fluid interior (see Kraus, 1967), then a simple relationship exists in the spectral domain between the load on the lithosphere and its deflection (see Banerdt, 1986; Beuthe, 2008; Turcotte et al., 1981; Willemann and Turcotte, 1981). If loading at only a single interface is considered (either at or below the surface), then the transfer function in eqn [40] is isotropic (i.e., independent of m). For this situation, expressions for the admittance and correlation functions can be written schematically as

$$Z(l) = f(\rho_c, \rho_m, v, E, T_e, T_c, z, g, R) \qquad [41]$$

$$\gamma(l) = 1 \quad \text{or} \quad -1 \qquad [42]$$

where f denotes a functional dependence on the enclosed parameters. In particular, the admittance function depends explicitly on the crustal and mantle density, Poisson's ratio v, Young's modulus E, the elastic thickness T_e, the crustal thickness, the depth of the load z, the magnitude of the gravitational acceleration g, and the radius of the planet. For an isotropic transfer function Q_l, it is trivial to show that the degree correlation function (in the absence of noise) is equal to either $+1$ or -1, according to the sign of Q_l. This model has been amended by McGovern et al. (2002), Belleguic et al. (2005), and Grott and Wieczorek (2012) to include in-phase loads applied to and below the surface when the two are related linearly by a degree-independent constant. Such models would include an additional parameter $\mathcal{L}$, which is a function of the relative magnitudes of the surface and subsurface loads. Geologic situations where surface and subsurface loads might be perfectly correlated include isolated volcanoes and impact basins.

An alternative loading model that includes loads applied to and below the surface was developed by Forsyth (1985) in the Cartesian domain (see also Banks et al., 2001). In contrast to models that take the applied loads to be perfectly in phase, it was assumed that the phase differences of the applied surface and subsurface loads were random. Such an assumption might be expected to be reasonable for continental cratons where several geologic processes have operated over extended periods of time (such as erosion, sedimentation, and magmatism). In contrast to eqn [42], this model possesses a wavelength-dependent correlation function.

A general model that allows for an arbitrary phase relationship between the surface and subsurface loads has been developed by the author in spherical coordinates and can be schematically described by the following equations:

$$Z(l) = f(\rho_c, \rho_m, v, E, T_e, T_c, z, \mathcal{L}, \alpha_l, g, R) \qquad [43]$$

$$\gamma(l) = f'(\rho_c, \rho_m, v, E, T_e, T_c, z, \mathcal{L}, \alpha_l, g, R) \qquad [44]$$

where both f and f' represent generic functional dependencies. The phase relationship of the two loads is described by an additional parameter α that can possess values between 1 and -1 for correlated and anticorrelated loads, respectively. The expectation of α is given by the expression

$$\alpha_l = \frac{\sum_{m=-l}^{l} \langle \cos \Delta_{lm} \rangle}{(2l+1)} \qquad [45]$$

where Δ_{lm} denotes the phase difference between the two loads and $\langle \cdots \rangle$ is the expectation operator. For random phases, α is zero, and the model degenerates to that of Forsyth (1985). When the loads are perfectly in or out of phase by 0° or 180°, $\alpha = \pm 1$ and the model is analogous to that of McGovern et al. (2002) and Belleguic et al. (2005). A similar model was developed contemporaneously by Kirby and Swain (2009) in Cartesian geometry using a somewhat different approach and then later by Audet (2013) in spherical harmonics.

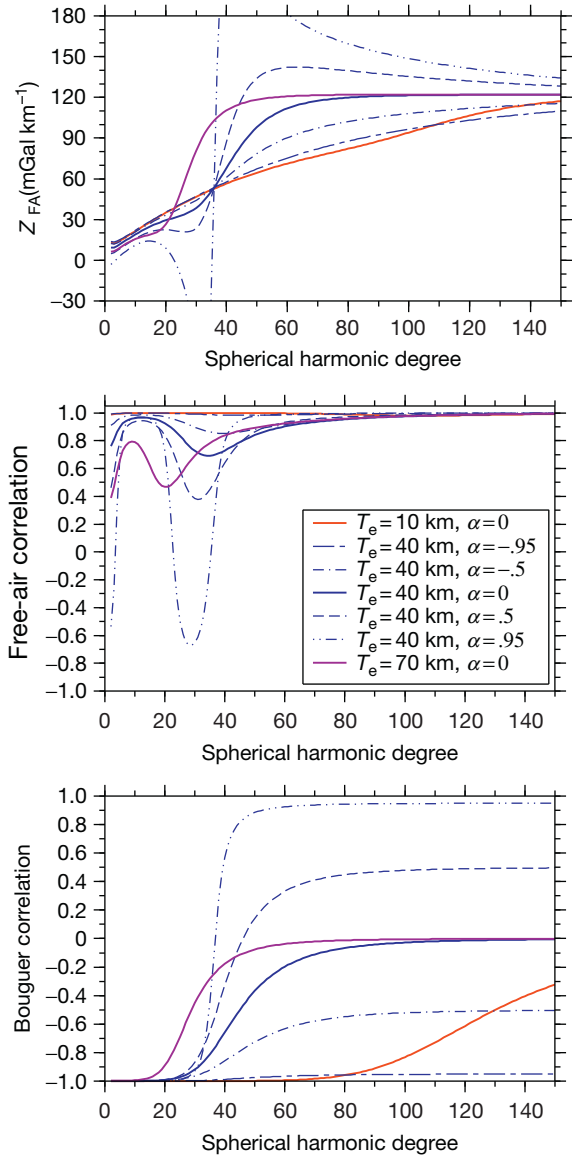


Figure 14 (Top) Free-air admittance, (middle) free-air correlation, and (bottom) Bouguer correlation, for a flexure model with equal magnitudes of applied surface and subsurface loads. Model parameters correspond to the planet Mars, with $T_c = z = 50$ km, $\rho_c = 2900$ kg m^{-3}, $\rho_m = 3500$ kg m^{-3}, and $E = 10^{11}$ Pa. The solid lines correspond to the case where the applied surface and subsurface loads have random phases (i.e., $\alpha = 0$), and the dashed lines correspond to the case where these loads are partially correlated.

Examples of the predicted free-air admittance and correlation functions are shown in Figure 14 for several values of the elastic thickness and phase parameter α. These models were generated using parameters typical for the planet Mars, and the magnitudes of the applied surface and subsurface loads were chosen to be equal. As is seen, these curves are strong functions of both the elastic thickness and α, and by considering both the admittance and correlation, it may be possible to separate the effects of the two. The free-air correlation function possesses low values over a restrictive range of wavelengths that is diagnostic of the elastic thickness. Furthermore, the free-air correlation function is seen to approach unity at high degrees ($l \gtrsim 100$). In practice, if a decline of the free-air correlation is observed with increasing degree, this is usually a good indicator of a loss of fidelity with the employed gravitational model. The predicted Bouguer correlation function is also shown for the same model parameters, and this shows a behavior similar to that predicted by the model of Forsyth (1985). It is noted that the Bouguer correlation asymptotically approaches the value of α with increasing degree. Whereas the Bouguer correlation is useful for interpretive purposes, its use is not advocated here because the Bouguer gravitational anomaly depends critically upon the value chosen for the crustal density, and this is in general not known a priori.

The preceding discussion assumed explicitly that the unmodeled gravitational signal I_{lm} in eqn [40] was identically zero. As mentioned in **Section 10.05.2.2**, if I_{lm} and the topography are uncorrelated, the expectation of the admittance will not be affected by the presence of such a signal. However, the expectation of the gravitational power spectrum will be biased upward by an additive constant S_{II}, and this will tend to bias the correlation function downward. In particular, the observed correlation function in the presence of noise is equal to

$$\gamma^{obs}(l) = \frac{\gamma(l)}{\sqrt{1 + S_{II}(l)/S_{gg}(l)}} \qquad [46]$$

where γ is the noise-free value.

Fortunately, any theoretical loading model can be easily modified to incorporate an unmodeled signal I_{lm} that is uncorrelated with the topography. It would be a simple matter to include S_{II} as an inversion parameter in a generic loading model such as eqn [44], though it is noted that this has not yet been attempted in the literature. While I_{lm} is commonly assumed to be gravitational measurement noise, this need not be the case. As described by McKenzie (2003), massive erosion or sedimentation of a lithosphere possessing surface and subsurface loads could give rise to a final state where the surface is perfectly flat but where gravitational anomalies are present. If this final state were taken as the initial condition of a *subsequent* loading event, then the initial gravitational signature (which is unrelated to the second loading model) would be expected to be uncorrelated with the subsequently generated topography.

Finally, it is important to reemphasize that if a given model of lithospheric loading is an accurate description of reality, it must fit both the admittance and coherence functions. If this cannot be done, then either the model or data must be inaccurate. Unfortunately, the vast majority of published investigations that use Forsyth-like loading models employ only the admittance or coherence function. The values of inverted parameters from such studies, while perhaps correct, need confirmation by analysis of the other function. Notable exceptions include the papers by Forsyth and coworkers (Bechtel et al., 1990, 1987; Ebinger et al., 1989; Forsyth, 1985; Pérez-Gussinyé et al., 2004; Phillips, 1994; Zuber et al., 1989). Similarly, many published investigations that employ a loading model with only surface or subsurface loads also ignore the correlation function, even though such models explicitly require this to be ± 1 when I_{lm} is zero.

10.05.7 Localized Spectral Analysis

The spherical harmonic functions represent a global basis, and the power spectrum as defined by eqn [9] is necessarily representative of the global properties of the function. In geophysics, however, it is reasonable to suspect that the spectral properties of the gravity and topography will vary as a function of position. For example, the elastic thickness may differ among geologic provinces as a result of their unique histories. Alternatively, it might arise that the data are known only locally and that one would like to estimate the power spectrum based exclusively upon these data. In this section, two methods are discussed that address these issues: One uses orthogonal windowing functions to form a multitaper spectral estimate, and the other uses spherical wavelets.

10.05.7.1 Multitaper Spectral Analysis

One way of obtaining localized estimates of a function's power spectrum is to first multiply the data by a localizing window (commonly referred to as a data taper) and then expand this windowed function in spherical harmonics (for a detailed discussion in the Cartesian domain, see Percival and Walden, 1993). However, as a result of the windowing procedure, the resultant power spectrum will differ from the true value. For the case where the input field is stationary and the spherical harmonic coefficients are governed by a zero-mean stochastic process, it can be shown that the expectation of the windowed power spectrum is related to the global spectrum by the following relation (Wieczorek and Simons, 2005, 2007):

$$\langle S_{\Phi\Gamma}(l)\rangle = \sum_{j=0}^{L} S_{hh}(j) \sum_{i=|l-j|}^{l+j} S_{fg}(i)\left(C^{l0}_{i0j0}\right)^2 \qquad [47]$$

Here, h represents an arbitrary window bandlimited to degree L, Φ, and Γ are the windowed fields hf and hg, respectively, and the last symbol in parentheses is a Clebsch–Gordan coefficient (these are related to the Wigner $3-j$ symbols and are proportional to the integral of three Legendre functions; see Varshalovich et al., 1988). The expectation of the windowed power spectrum $S_{\Phi\Gamma}$ is seen to be related to the global spectrum by a smoothing operation reminiscent of a convolution.

For a localized spectral analysis, the question arises naturally as to what form the localizing window should take. In order to spatially localize the data, it is clear that the amplitude of the window (or its power) should be near zero outside the region of interest. Furthermore, as a result of the convolution-like relationship between the global and windowed spectra, the bandwidth L of the window should be as small as possible in order to limit this spectral smoothing. Slepian and coworkers (see Slepian, 1983) posed and solved this problem in Cartesian geometry by finding those windows whose power was concentrated optimally in a specified region. Using this same criterion, Wieczorek and Simons (2005) and Simons et al. (2006) solved for those bandlimited windows on the sphere that are concentrated optimally within a spherical cap. The case of arbitrary regions was treated by Simons et al. (2006), and the case where data are absent at the poles (the polar gap problem) was treated by Simons and Dahlen (2006).

This optimization problem reduces to a simple eigenvalue equation whose solution yields a family of $(L+1)^2$ orthogonal data tapers; the quality of the concentration is given by the corresponding eigenvalue, λ, which is the fractional power of the function within the region of interest. As an example, for a concentration region specified by a spherical cap with an angular radius of 30°, as shown in **Figure 15**, 34 windows exist with a bandwidth $L=29$ that concentrate more than 99% of their power within the region of interest.

Though it is common to use a single localization window when performing a localized spectral analysis, the use of multiple tapers, as pioneered by Thomson (1982) in the Cartesian domain, can be highly advantageous. The multitaper estimate is defined as the weighted sum of spectra obtained from using K well-localized orthogonal tapers:

$$S^{(mt)}_{\Phi\Phi}(l) = \sum_{k=1}^{K} a_k S^{(k)}_{\Phi\Phi}(l) \qquad [48]$$

The values of the weights are usually assigned to be either $1/K$ or proportional to $1/\lambda_k$, but these can be chosen to also minimize the variance of the estimate (Wieczorek and Simons, 2007). The benefits of using a multitaper estimate are several fold. First, the cumulative energy of orthogonal tapers will be nearly constant across the region of interest, whereas the energy of any single window will nonuniformly cover the concentration region. Because of this, a multitaper average will be more representative of the data than that of a single taper. Second, while it is generally not possible to obtain the expectation of the localized spectrum since there is usually only a single field available for analysis, the spectral estimates obtained from orthogonal tapers are nearly uncorrelated, and their average can be considered as an approximation of the expectation. Finally, by using multiple tapers, it is possible to make an estimate of the uncertainty associated with a given spectral estimate; this is expected to decrease as approximately $1/\sqrt{K}$ (Wieczorek and Simons, 2005, 2007).

In performing a localized spectral analysis, there are several factors that need to be considered. First, it is noted that if the fields f and g of eqn [47] are known to a maximum spherical harmonic degree of L_{fg}, then only the first $L_{fg}-L$ windowed spectral estimates are reliable. Second, those localized spectral estimates with degrees less than L are heavily biased and possess relatively large uncertainties, making these of little use for geophysical analysis. Third, while a multitaper spectral analysis is generally preferable to using a single taper, the aforementioned two concerns present serious limitations when working with the relatively low-resolution gravity fields of Mercury, Venus, and Mars. Depending on the size of the concentration region, it may be infeasible to use the larger bandwidths that are required for obtaining several well-concentrated tapers.

Finally, when comparing model results to observations, it is emphasized that the two must be windowed in the same manner (e.g., Pérez-Gussinyé et al., 2004). If the analysis is performed by generating forward models of the gravity field using the known topography, then it is necessary to localize these functions in the same manner as the data. Alternatively, if no explicit expression exists for Q_{lm} (as in the model of Forsyth (1985) and that presented in **Section 10.05.6.2**, both of which are statistical in nature), then it is necessary to window the

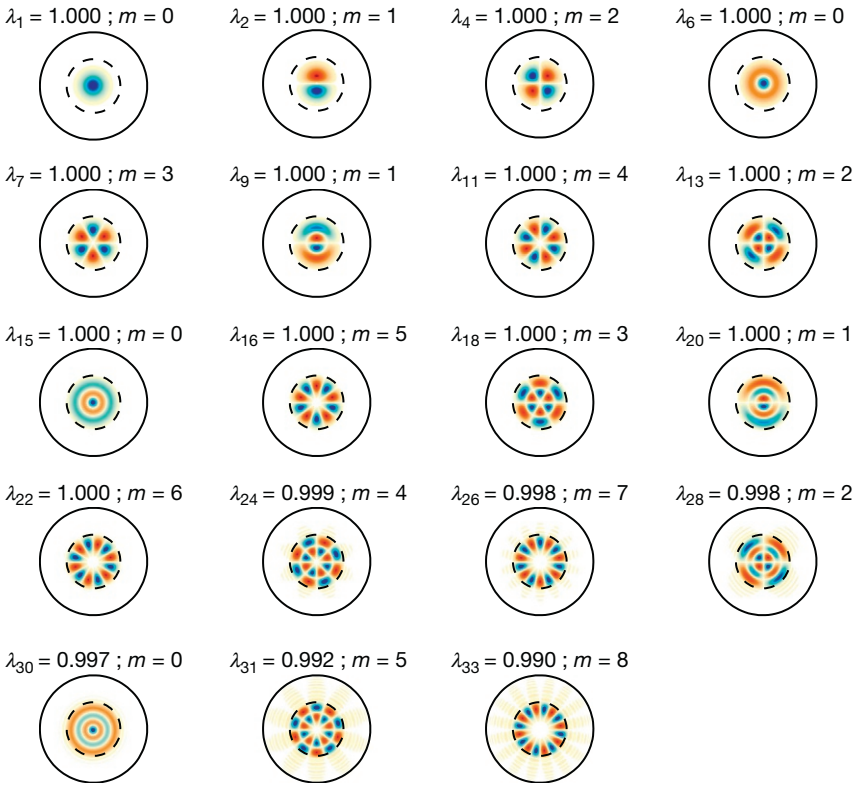

$\lambda_1 = 1.000$; $m = 0$
$\lambda_2 = 1.000$; $m = 1$
$\lambda_4 = 1.000$; $m = 2$
$\lambda_6 = 1.000$; $m = 0$
$\lambda_7 = 1.000$; $m = 3$
$\lambda_9 = 1.000$; $m = 1$
$\lambda_{11} = 1.000$; $m = 4$
$\lambda_{13} = 1.000$; $m = 2$
$\lambda_{15} = 1.000$; $m = 0$
$\lambda_{16} = 1.000$; $m = 5$
$\lambda_{18} = 1.000$; $m = 3$
$\lambda_{20} = 1.000$; $m = 1$
$\lambda_{22} = 1.000$; $m = 6$
$\lambda_{24} = 0.999$; $m = 4$
$\lambda_{26} = 0.998$; $m = 7$
$\lambda_{28} = 0.998$; $m = 2$
$\lambda_{30} = 0.997$; $m = 0$
$\lambda_{31} = 0.992$; $m = 5$
$\lambda_{33} = 0.990$; $m = 8$

Figure 15 Spatial rendition of those functions that are near-perfectly concentrated within a spherical cap of angular radius 30° with a spherical harmonic bandwidth $L = 29$. The dotted circle corresponds to the localization region, and the blue and red colors correspond to negative and positive values, respectively. For these parameters, the total number of functions with $\lambda > 0.99$ is equal to 34. Each nonzonal function has a twin that is rotated by 90°/m. Image reprinted from Wieczorek MA and Simons FJ (2007) Minimum-variance multitaper spectral estimation on the sphere. *Journal of Fourier Analysis and Applications* 13: 665–692, http://dx.doi.org/10.1007/s00041-006-6904-1 with permission from Springer Science and Business Media.

predicted (cross) power spectra using eqn [47] before calculating the theoretical admittance and correlation functions.

10.05.7.2 Wavelet Analysis

An alternative approach of performing a localized spectral analysis is to use spherical wavelets. A wavelet analysis differs from a multitaper analysis in that (1) the function is multiplied by a wavelet of a given size and integrated over all space, as opposed to being multiplied and then expanded in spherical harmonics, and (2) the wavelet scalogram (the analog of the power spectrum) is a function of the wavelet scale (i.e., size) and not spherical harmonic degree. Whereas the multitaper spectral analysis approach is best when the function is known to be locally stationary within the analysis region, wavelet analyses are better suited to truly nonstationary processes.

The wavelet transform for real functions is defined as the inner product of a function f with a wavelet ψ of scale s centered at position $\hat{\mathbf{r}}$ and with angular orientation γ about this axis:

$$\tilde{f}(s, \hat{\mathbf{r}}, \gamma) = \int_\Omega f(\theta', \phi') \psi_{s,\hat{\mathbf{r}},\gamma}(\theta', \phi') \mathrm{d}\Omega' \qquad [49]$$

The wavelet transform can be defined either to be continuous in scale s or to be composed of discreet scales. By expressing both f and ψ in spherical harmonics,

$$\tilde{f}(s, \hat{\mathbf{r}}, \gamma) = 4\pi \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} \psi_{lm}(s, \hat{\mathbf{r}}, \gamma) \qquad [50]$$

where L is the smaller of the spherical harmonic bandwidth of f and ψ (the product is zero beyond the bandwidth of either function). If the wavelets are zonal functions, the wavelet transform is

$$\tilde{f}(s, \hat{\mathbf{r}}) = 4\pi \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} \Psi_l(s) \sqrt{\frac{1}{2l+1}} Y_{lm}(\theta, \phi) \qquad [51]$$

where Ψ_l are the spherical harmonic coefficients of the zonal wavelet centered at the north pole. For zonal wavelets, the wavelet transform is equivalent to the spatial convolution of f and ψ. The wavelet transform can be calculated rapidly everywhere using fast Fourier transform techniques and Wigner rotation matrices (Wandelt and Gòrski, 2001).

The wavelet cross scalogram of two real functions f and g is defined as

$$P_{fg}(s, \hat{\mathbf{r}}, \gamma) = \tilde{f}(s, \hat{\mathbf{r}}, \gamma) \tilde{g}(s, \hat{\mathbf{r}}, \gamma) \qquad [52]$$

If it is assumed that the two functions f and g are stochastic stationary processes, it can be shown that the expectation of this expression is related to the power spectrum of the wavelet and the cross power spectrum of f and g by

$$\langle P_{fg}(s)\rangle = (4\pi)^2 \sum_{l=0}^{L} \frac{S_{fg}(l) S_{\Psi_s\Psi_s}(l)}{2l+1} \quad [53]$$

where $S_{\Psi_s\Psi_s}$ is the power spectrum of the wavelet of scale s. This fundamental relation is analogous to eqn [47] as used in localized spectral analyses and quantifies how the power spectrum of the data is modified by the window spectrum when forming the wavelet cross scalogram.

In general, wavelets should have a zero mean, they should be localized in space in order to study nonstationary processes, and they should be localized in the spectral domain in order to best extract the frequency content of the signal (cf. eqn [53]). A multitude of wavelet families have been constructed, including those that are zonal (Chambodut et al., 2005; Fengler et al., 2007; Freeden and Windheuser, 1997; Guilloux et al., 2009; Holschneider et al., 2003; Holschneider and Iglewska-Nowak, 2007; Li, 1999; Michel and Wolf, 2008; Panet et al., 2011; Schmidt et al., 2007; Wiaux et al., 2007) and directional (Audet, 2011, 2013; McEwen et al., 2007). Once the family of wavelets has been constructed for each scale s, it is straightforward to construct the wavelet transform. Nevertheless, the best manner to construct a family of wavelets on the sphere is not trivial. The biggest problems consist of how to discretize the wavelets as a function of scale, how to most optimally represent the data by the smallest number of wavelet coefficients as possible, and how to minimize correlations between the wavelet coefficients (e.g., Freeden and Windheuser, 1997; Narcowich and Ward, 1996). A perhaps more daunting problem is deciphering the spherical wavelet literature itself: many of these papers are highly mathematical, the notation is highly variable among authors, and the papers span a wide range of disciplines and journals.

10.05.8 Summary of Major Results

10.05.8.1 Earth

The gravity and topography of Earth have been used extensively to decipher the rheological properties of the crust and upper mantle. The literature is voluminous, and the reader is referred to several reviews in volume 8 of this series, Watts (2001), and the references in the papers cited in the succeeding text. Here, only a few subjects will be touched upon that bear relevance to investigations of Venus, Mars, Mercury, and the Moon. These include modeling of the elastic thickness of the oceanic and continental lithosphere, inelastic flexural modeling, and the modeling of dynamic topography and geoid signatures associated with mantle convection.

Flexural modeling of the oceanic lithosphere is relatively simple in that loading is primarily a result of the construction of isolated shield volcanoes. Elastic thickness estimates have been obtained by modeling the topographic and gravity signatures of these features, and it is widely accepted that the elastic thickness is dependent primarily upon the age of the plate at the time of loading, with T_e being less than about 45 km (for a review, see Watts, 2001). In particular, a plot of the elastic thickness versus age of the lithosphere at the time of loading resembles the time dependence of the depth to an isotherm (~300–600 °C) predicted from a plate cooling model (see Watts and Zhong, 2000). This suggests that the flexural signature has been 'frozen' into the lithosphere as it cooled and that subsequent long-term viscoelastic relaxation has been relatively unimportant. Nevertheless, a description of the initial short-term subsidence of the lithosphere (i.e., the first few 10 s of ka) requires the use of a viscoelastic model, and given the relatively restricted age range of oceanic lithosphere (<200 Ma), it is difficult to discern if viscoelastic relaxation would be important at longer timescales. It is important to note that most flexural modeling of features on the other terrestrial planets has been for loads that were emplaced on the lithosphere over a billion years ago.

Investigations of the continental elastic thickness have been more contentious. Part of the difficulty arises because the importance of subsurface loading, and the phase relationship of the surface and subsurface loads, is not known a priori (see **Section 10.05.6.2**). A loading model was developed by Forsyth (1985) that took into account both surface loading and subsurface loading under the assumption that the two were uncorrelated, and application of this method has yielded elastic thicknesses in the broad range from about 5 to 150 km (Bechtel et al., 1990, 1987; Ebinger et al., 1989; Forsyth, 1985; Kirby and Swain, 2009; Pérez-Gussinyé et al., 2004, 2009; Zuber et al., 1989). There is some controversy as to whether the values greater than ~25 km are reliable (compare McKenzie, 2003 with Watts and Burov, 2003), but this debate will not be settled definitively until investigators fit both the admittance and correlation functions simultaneously while taking into account partially correlated loads and unmodeled gravitational signals (see **Section 10.05.6.2**). Furthermore, some studies that have inverted for the elastic thickness using multitaper spectral analysis and wavelet techniques have done so using a methodology that is not entirely correct. In particular, as both windowing and wavelet approaches return a modified form of the power spectrum (cf. eqns [47] and [53]), it is necessary to compare these modified spectra with suitably modified theoretical spectra, which is not always done. Regardless, these methods have convincingly shown that the elastic thickness of some continental regions is not always isostropic (e.g., Kirby and Swain, 2006; Simons et al., 2003, 2000), which is an assumption common to most studies.

The majority of investigations that model lithospheric flexure assume that the lithosphere is perfectly elastic, but elastic stresses are often predicted to be in excess of the strength of geologic materials. A simple modification to the elastic flexure equation that takes this into account is to replace the elastic bending moment–curvature relationship with one that is based upon an elastic–perfectly plastic (EP) model of the lithosphere's yield strength (e.g., Burov and Diament, 1995; Mueller and Phillips, 1995). In particular, the strength of the upper crust is limited by brittle failure, and stresses in the lower crust and mantle are limited by their ductile strength for a specified strain rate. Predicted flexural profiles are time-invariant and can sometimes differ significantly from those of the perfectly elastic model. As the ductile strength is temperature-dependent, these results are sensitive to the assumed lithospheric temperature gradient.

A more realistic model of lithospheric deformation uses a time-dependent elastoviscoplastic (EVP) formulation (e.g., Albert and Phillips, 2000; Albert et al., 2000; Brown and

Phillips, 2000). The main advantage of these models is that the strain rates are calculated explicitly, as opposed to assumed as in the EP models. Though the best-fit EP and EVP flexural profiles can be quite similar, it is not clear a priori how one should estimate the characteristic strain rate that is required for the EP model without running a full EVP simulation (Albert et al., 2000). The EVP models show that significant decoupling of stresses may occur between the crust and mantle if the lower crust is sufficiently weak (e.g., Brown and Phillips, 2000). When this occurs, the effective elastic thickness decreases; the exact value is highly dependent upon the crustal thickness, load magnitude, and assumed rheology of the crust and mantle. In contrast, when the lower crust is strong, the maximum achievable effective elastic thicknesses are consistent with the depth of an ~700 °C isotherm obtained from a lithospheric cooling model. Flexural modeling of a volcano growing on a cooling lithosphere shows that the effective elastic thickness is 'frozen' into the lithosphere shortly after volcanic construction is complete (Albert and Phillips, 2000).

Finally, in addition to near-surface crustal thickness and density variations, significant gravity and topography signatures can be generated by dynamic processes in the mantle, such as beneath hot spots and subduction zones. Though there are few, if any, convincing examples of plate subduction on the other terrestrial planets, hot spots similar to Earth are believed to exist on both Venus and Mars. Dynamic modeling of plumes shows that the major variable controlling the surface gravity and topography signatures is the depth dependence of the mantle viscosity. In the absence of a shallow low-viscosity asthenosphere, convective stresses generated at depth are coupled efficiently to the surface, generating large signals and large corresponding effective depths of compensation. However, the inclusion of a shallow low-viscosity zone can reduce these signatures significantly, giving rise to shallower apparent depths of compensation (e.g., Ceuleneer et al., 1988; Robinson and Parsons, 1988). Joint inversions utilizing mantle density anomalies from seismic tomography and estimates for the dynamic topography signal imply the existence of a low-viscosity zone somewhere in the upper mantle and a gradual increase in viscosity with depth by an order of magnitude in the lower mantle (e.g., Panasyuk and Hager, 2000).

10.05.8.2 Venus

Our knowledge of Venus has dramatically improved since the acquisition of gravity, topography, and SAR imagery by the Magellan mission between 1990 and 1994. The size and bulk density of Venus were known beforehand to be similar to that of Earth, but this planet was found to differ dramatically in that it lacks any clear sign of plate tectonics. A major unanswered question is how this planet loses its internal heat and whether or not this process is episodic or uniform in time. Geophysical analyses have been used to constrain the crustal and elastic thickness, and the latter has been used to place constraints on the temperature gradient within the lithosphere. Reviews concerning the geophysics of this planet can be found in Phillips et al. (1997), Grimm and Hess (1997), Schubert et al. (1997), and Nimmo and McKenzie (1998).

The crustal plateaus of Venus have generally low-amplitude gravitational and topographic signatures within their interiors and low GTRs. With the exception of Ishtar Terra, these regions are potential candidates for being isostatically compensated. Assuming that the surface topography is compensated at a single interface, Grimm (1994) obtained best-fit apparent depths of compensation between 20 and 50 km for Alpha, Tellus, Ovda, and Thetis Regiones. These values are plausibly interpreted as representing the depth of the crust–mantle interface. Analyses of GTRs by Kucinskas and Turcotte (1994), Moore and Schubert (1997), and James et al. (2013) obtained zero-elevation crustal thicknesses between about 40 and 80 km for Alpha, Ovda, Tellus, and Thetis Regiones. If any portion of the geoid and topography were to result from Pratt or thermal compensation, the obtained crustal thicknesses would represent an upper bound. A spectral admittance study by Phillips et al. (1997) obtained a slightly thinner crustal thickness of 38 ± 9 km for the region of Atla Regio (see also Phillips, 1994).

In contrast to the majority of the highland plateaus, large apparent depths of compensation and GTRs have been found for the volcanic rises (Kucinskas and Turcotte, 1994; Moore and Schubert, 1995, 1997; Smrekar, 1994) and Ishtar Terra (Grimm and Phillips, 1991; Hansen and Phillips, 1995; Moore and Schubert, 1997). These values are not consistent with compensation occurring solely by crustal thickening, but require some form of dynamic support from the mantle via stresses induced by ascending mantle plumes (e.g., Vezolainen et al., 2004) or substantial thinning of a thick (~300–600 km) thermal lithosphere (Kucinskas and Turcotte, 1994; Moore and Schubert, 1995, 1997; Orth and Solomatov, 2011). If a low-viscosity asthenosphere were present at shallow mantle depths, as is the case of Earth, the predicted GTRs and apparent depths of compensation resulting from dynamic support would be considerably smaller than measured as a result of the decoupling of stress between the lithosphere and mantle (e.g., Kiefer and Hager, 1991; Kiefer et al., 1986). This seems to imply that in contrast to Earth, Venus lacks a low-viscosity zone (Huang et al., 2013; Pauer et al., 2006), which is most likely a result of a volatile-poor mantle. Modeling by Pauer et al. (2006) suggests that the viscosity of the Venusian mantle could increase by a factor of about 10–80 from the upper to lower mantle, similar to that of Earth. A strong coupling of stress between the lithosphere and convecting mantle is the likely cause of the high spectral correlation between gravity and topography for the lowest spherical harmonic degrees of Venus (contrast **Figures 2** and **4**).

An additional constraint concerning the thickness of the Venusian crust is related to its compositional buoyancy. If the crust were basaltic in composition, this material would undergo a phase transition at high pressure to the more dense mineral assemblage of eclogite (e.g., Grimm and Hess, 1997). This material could potentially delaminate from the base of the crust as a result of its high density, and the depth of this phase transition might thus constrain the maximum achievable crustal thickness. For a MORB composition, the eclogite phase transition is predicted to occur at depths of ~70–120 km (see Ghent et al., 2004) for linear temperature gradients of 5 and 15 K km^{-1}, respectively.

The crustal thickness model presented in **Figure 12** has thicknesses near 70 km only in the highland plateaus of Ishtar Terra and Ovda and Thetis Regiones, suggesting that crustal delamination could have occurred in these regions. The

highland crust is almost exactly twice as thick as the lowland crust, which could potentially arise by the obduction of one plate over another. In contrast to Mercury, Mars, and the Moon, no large impact basins are found on this planet as a result of the relatively young age of its surface (~240 Ma; Le Feuvre and Wieczorek, 2011). This crustal thickness model assumed an average thickness of 35 km and did not consider density anomalies beneath the crust–mantle interface. An alternative model was developed by Orth and Solomatov (2012) who assumed that variations in lithospheric thickness were isostatically compensated and that the remaining signal was a result of isostatic variations in crustal thickness. James et al. (2013), in contrast, constructed a global crustal thickness model by assuming that the observed surface topography and gravitational potential were a result of isostatically compensated crust and dynamic relief arising from mantle density anomalies below an elastic lid.

Elastic thickness estimates have been obtained for a variety of features based exclusively on topographic profiles that are indicative of flexure. The benefit of using topography alone is that small features can be investigated that are not resolved in the current gravity model. Elastic thicknesses of 11–25 km have been obtained by modeling the Freyja Montes foredeep (Sandwell and Schubert, 1992; Solomon and Head, 1990) and ~10–60 km for potential flexural bulges outboard of coronae (Sandwell and Schubert, 1992). Additional features modeled by Johnson and Sandwell (1994) yield elastic thicknesses of 10–40 km, and potential subduction-related sites possess a range of 6–45 km (Schubert and Sandwell, 1995). Predicted stresses are largest where the plate curvature is greatest, and faulting is generally visible in the Magellan SAR imagery at these locations. Modeling by Barnett et al. (2002) yielded best-fit elastic thicknesses that are consistent with the aforementioned studies. Modeling the location of concentric faulting around Nyx Mons (a volcano in Bell Regio) implies a best-fit elastic thickness of ~50 km (Rogers and Zuber, 1998). It is important to note that the assumption of a perfectly elastic rheology may be grossly inappropriate for some features. For instance, the magnitude of the flexure-induced bulge south of Artemis Chasma implies that significant yielding has occurred within the lithosphere, and inelastic modeling by Brown and Grimm (1996) shows that a significant compressive in-plane force is required at this locale.

The elastic thickness has also been estimated for various regions of Venus through a combined analysis of gravity and topography data in the spectral domain. However, the majority of these investigations have fit only the admittance or coherence functions, but not both simultaneously (e.g., Anderson and Smrekar, 2006; Barnett et al., 2000, 2002; Hoogenboom et al., 2005, 2004; Lawrence and Phillips, 2003; McKenzie, 1994; McKenzie and Nimmo, 1997; Simons et al., 1994, 1997; Smrekar, 1994; Smrekar et al., 2003; Smrekar and Stofan, 1999). The robustness of the inverted parameter values, the fidelity of the gravity model, and the appropriateness of the loading model are all difficult to assess in these studies. One exception is that of Phillips (1994) who investigated the lithospheric properties of Atla Regio, which is believed to be an active hot spot based on its geomorphology and previously determined large apparent depths of compensation. Using the loading model of Forsyth (1985), which assumes uncorrelated surface and subsurface loads, it was shown that a single mode of compensation could not explain the entire wavelength range of the admittance and coherence functions. An inversion utilizing only the short wavelengths yielded a crustal thickness of 30 ± 13 km and an elastic thickness of 45 ± 3 km. Whereas surface loading by the volcanic constructs in this area dominates, about 10% of the load is required to be located at shallow depths within the crust. For the long-wavelength range, only the depth of the subsurface load was well constrained with a value near 150 km, and the elastic thickness was constrained only to be less than 140 km.

It is possible to place constraints on the crustal temperature gradient at the time of loading either by using the obtained elastic thicknesses or by forward modeling of inelastic flexure. The basic approach is to match the bending moment implied by the elastic model to that predicted by an inelastic rheology (McNutt, 1984). The obtained temperature gradient estimates lie in the rather broad range of 3–26 K km^{-1} (Brown and Grimm, 1996; Johnson and Sandwell, 1994; Phillips, 1994; Phillips et al., 1997; Sandwell and Schubert, 1992), but the majority of these lie on the low end, between ~4 and 10 K km^{-1}. This is considerably lower than the expected Earth-scaled temperature gradient of ~15 K km^{-1} (e.g., Phillips, 1994), especially when considering that some of these estimates were obtained where an underlying mantle plume might be expected. Though such calculations are critically dependent on the validity of the inelastic strength model, the assumed strain rate, and the depth dependence of temperature, these results seem to imply that the background heat flow of Venus is much less than would be expected by analogy to Earth. Such an interpretation is consistent with a model in which the Venusian lithosphere formed catastrophically ~500–1000 Mya and has since been cooling conductively (cf. Moresi and Solomatov, 1998; Parmetier and Hess, 1992; Turcotte, 1995). However, such a model is not required, or even preferred, by the cratering history of the Venusian plains (Hauck et al., 1998).

10.05.8.3 Mars

Following the acquisition of high-resolution gravity and topography data from the MGS and Mars Odyssey missions, a number of studies have been published bearing on the crustal and lithospheric structure of Mars. These investigations have placed constraints on the thickness of the Martian crust, the elastic thickness, and the crustal density and also imply the existence of dynamic support of topography and buried mass anomalies. Reviews concerning the gravity, topography, and crust of Mars can be found in Esposito et al. (1992), Banerdt et al. (1992), Zuber (2001), Wieczorek and Zuber (2004), Nimmo and Tanaka (2005), and Grott et al. (2013).

The average thickness of the Martian crust has been constrained by the analysis of GTRs over the ancient southern highlands. After removing the long-wavelength flexural and load signatures associated with the Tharsis province, a zero-elevation thickness of 57 ± 24 km was obtained under the assumption of Airy isostasy (Wieczorek and Zuber, 2004). This range of values is consistent with estimates based upon the viscous relaxation of topography (Nimmo and Stevenson, 2001; Zuber et al., 2000b) and geochemical mass-balance arguments, both of which require the crust to be less than ~100 km thick (see Wieczorek and Zuber, 2004). Crustal

thickness modeling further requires the mean thickness of the crust to be greater than about 25 km (Pauer and Breuer, 2008).

A global model of the crustal thickness of Mars has been constructed by Neumann et al. (2004) by assuming a mean thickness of 45 km (see **Figure 12**). If the assumption of a constant density crust is correct, the crust of the southern highlands is predicted to be thicker by about 30 km than the northern lowlands. However, if the northern lowland crust is denser than the southern highlands, as implied by the results of Belleguic et al. (2005), then the actual crustal thickness difference could be significantly less. The Tharsis province is seen to possess a relatively thick crust, indicative of prolonged volcanic construction, whereas the crust beneath the major impact basins is thinned and in some places nearly absent.

Localized spectral admittance and correlation spectra have been modeled in spherical coordinates for various regions using the techniques of Simons et al. (1997) and Wieczorek and Simons (2005). In the investigations of McGovern et al. (2002, 2004), Belleguic et al. (2005), Wieczorek (2008), Grott and Wieczorek (2012), and Beuthe et al. (2012), a thin elastic spherical shell loading model was employed that depended upon the shell's elastic thickness, the load density, the crustal density, and the ratio of the magnitudes of subsurface and surface loads, which were assumed to be in or out of phase by 0° or 180°. When the load density differs from that of the crust, the methodology of Belleguic et al. (2005) is more accurate, and their results are here emphasized.

Of all the parameters considered by Belleguic et al. (2005) when modeling the major Martian volcanoes, the load density is in general the best constrained. For the Tharsis Montes and Elysium rise, values of $\sim 3200 \pm 100\ \mathrm{kg\ m^{-3}}$ were obtained. After accounting for a few-percent porosity, this range is consistent with density estimates of the Martian meteorites, which are thought to be derived from these regions based on their young ages. Elastic thickness estimates are somewhat variable but were found to lie between about 50 and 100 km when only surface loads were considered. However, when both surface and subsurface loads were modeled, only upper and lower bounds could be specified for most regions. The crustal density was constrained only beneath the Elysium rise (which is located in the northern lowlands) and was found to be identical to the density of the superposed load. Combined modeling of GTRs and crustal thickness suggest a maximum density of about $3020\ \mathrm{kg\ m^{-3}}$ for the southern highlands, which is consistent with the value of $3000\ \mathrm{kg\ m^{-3}}$ suggested by Neumann et al. (2004) based on rock compositions at the Mars Pathfinder site. If the crustal density beneath the Elysium rise is representative of the northern lowlands, then this implies a hemispheric dichotomy in crustal composition. This result is consistent with geochemical maps obtained from the Mars Odyssey gamma-ray spectrometer that show the northern lowlands possess a higher iron concentration than the southern highlands (Taylor et al., 2006). The low elevations of the northern plains could thus be a partial result of Pratt compensation. Finally, the inclusion of less dense subsurface loads (either compositional or thermal in origin) improved the misfit between the modeled and observed admittance functions.

In the localized admittance study of Beuthe et al. (2012), two-stage loading models were employed to investigate how the elastic thickness and load density varied with time. This loading model could fit the data and models with single-stage top and bottom loading, calling into question the necessity of subsurface loading. Using similar localized admittance techniques, density and elastic thickness have been estimated at other regions on the planet. For the isolated low-relief volcano Tyrrhena Patera, Grott and Wieczorek (2012) obtained a density between 2960 and $3500\ \mathrm{kg\ m^{-3}}$, which is consistent with the Tharsis and Elysium volcanoes, and an elastic thickness that is less than 28 km. Zuber et al. (2007) and Wieczorek (2008) investigated the ice-rich south polar layered deposits and found densities close to $1250\ \mathrm{kg\ m^{-3}}$.

Elastic thickness estimates for other locales have been determined using a variety of techniques, but these generally contain a larger number of assumptions. Modeling of the geologically inferred flexural moat of the northern polar cap suggests an elastic thickness between 60 and 120 km (Johnson et al., 2000). Faults surrounding the Isidis impact basin imply elastic thicknesses between 100 and 180 km at the time of loading (Ritzer and Hauck, 2009). If the topography of the dichotomy boundary is flexural in origin, with loading in the northern plains, then elastic thicknesses of ~31–36 km are implied (Watters, 2003). Cartesian admittances were analyzed in several other studies (Hoogenboom and Smrekar, 2006; McKenzie et al., 2002; Milbury et al., 2009; Nimmo, 2002), but the validity of the loading model was not quantified by calculating theoretical correlation functions. A summary of how elastic thickness constrains Martian heat flow can be found in Grott et al. (2013).

One distinctive feature of Mars is the large geoid and topography signals associated with the Tharsis province (e.g., Phillips et al., 2001). Two possible end-member explanations for this observation are either that it is a result of voluminous extrusive lavas that are partially supported by the lithosphere or that it is a result of dynamic topography associated with an underlying plume. Viscoelastic modeling of the geoid and topography response to internal buoyant loads implies that a plume is incapable of producing the totality of the observed signals (Roberts and Zhong, 2004; Zhong, 2002). By modeling the contributions of both surface and plume signals with a method that approximates a viscoelastic response, the degree-2 and degree-3 GTRs imply that a plume can account for only ~15% and 25% of the geoid and topography signals, respectively. Using a modified approach that includes all spherical harmonic degrees, Lowry and Zhong (2003) inverted for the relative contributions of surface and internal loads and found that a plume could account only for a maximum of ~25% and 50% of the observed geoid and topography, respectively.

Finally, it is noted that buried mass anomalies can be investigated by examination of the residual gravity field after subtraction of an appropriate reference model. By modeling the gravity field of the Syrtis Major region by surface loading of an elastic shell, an unmodeled localized density anomaly was found beneath this volcanic province. The amplitude of this unmodeled anomaly is consistent with the presence of dense cumulates of an extinct magma chamber (Kiefer, 2004). Using a similar technique, buried mass anomalies have also been inferred along the portions of the dichotomy boundary (Kiefer, 2005).

10.05.8.4 Mercury

Our understanding of Mercury is currently being revolutionized by the analysis of data from the MESSENGER spacecraft.

At the time of this writing, the initial analyses have just been published, and more result will soon follow.

Mercury is somewhat similar to the Moon in that its surface records billions of years of impact cratering. One striking aspect concerning the topographic expression of the largest impact basins is that they are not as topographically distinct as those found on the Moon and Mars. Many of these basins have shallower than expected depths (Mohit et al., 2009), and others have been deformed topographically. One clear example of basin deformation that postdates the impact event is the topographic bulge in the northern portion of the Caloris basin that is higher than most of its topographic rim (Zuber et al., 2012). Together, these observations suggest some form of modification by either mantle convection or crustal viscous relaxation (Mohit et al., 2009).

The Caloris impact basin is clearly associated with a positive gravitational anomaly, similar to the mascon basins on the Moon. Few other basins have corresponding gravitational anomalies, but this could be an artifact of the low resolution of the current gravity model. Crustal thickness modeling shows that the Caloris, Budh, and Sobkou basins are associated with a thinned crust (Smith et al., 2012). However, little is known about the average thickness of the crust, which is important for understanding the bulk silicate composition of the planet. Estimates of the depths of faulting constrain the depth of the brittle–ductile transition, and when combined with an estimate of the melting temperature of the base of the crust, an average crustal thickness less than 140 km is required (Nimmo and Watters, 2004).

The lithospheric thickness of Mercury appears to have varied over time. Many regions of ancient high-standing topography do not possess any significant gravitational anomaly, indicating that they were likely in a state of isostatic compensation when they formed. In contrast, a positive gravitational anomaly is associated with a domical rise in the northern lowlands, and modeling of this feature implies that the lithosphere had an elastic thickness of 70–90 km at the time of its formation (Smith et al., 2012). As a result of Mercury's eccentric orbit and 3/2 spin–orbit resonance, the average temperature of the surface has strong variations in both latitude and longitude, with two hot 'poles' being located at 0° and 180° on the equator. These temperature variations are predicted to give rise to long-wavelength variations in the thickness of the lithosphere, with variations in the mechanical thickness approaching 80 km between ±90°N and the equatorial hot poles (Williams et al., 2011).

10.05.8.5 The Moon

Analyses of the gravitational field of the Moon have been hindered historically by the lack of direct tracking data over the lunar farside hemisphere. This problem was rectified only recently by the collection of four-way tracking data from the Kaguya (2007) mission and intersatellite range-rate measurements from the Gravity Recovery and Interior Laboratory (2011) mission. Together with altimetric data and stereo imagery collected by the LRO and Kaguya missions, it can be said that the gravity and topography of the Moon are known better in a global sense than any other terrestrial object. These data have permitted a wide range of investigations, including the mapping of crustal thickness variations, studies of giant impact basins, and the mapping of subsurface density anomalies. A comprehensive review of results predating the GRAIL, LRO, and Kaguya missions can be found in Wieczorek et al. (2006).

The observed gravity signal at short wavelengths, where lithospheric flexure is unimportant, should be almost entirely a result of relief along the surface (cf. **Figure 14**). Wieczorek et al. (2013) used this observation to estimate the bulk density of the lunar highland crust. The obtained value of 2550 kg m^{-3} was substantially lower than the expected grain density of about 2900 kg m^{-3}, and this implies the existence of about 12% porosity in the upper crust. Although the depth dependence of porosity is not well constrained, it must extend at least several kilometers below the surface and perhaps into the upper mantle. These high porosities are almost certainly a result of fractures generated by billions of years of impact cratering and are consistent with impact-induced porosities associated with small impact craters on Earth (see also Huang and Wieczorek, 2012).

A global crustal thickness model of the Moon can be constructed by assuming values for the density of the crust, the density of the mantle, and the mean crustal thickness (see **Section 10.05.5**). The density of the lunar crust can be estimated from either remote sensing data or analyses of the short-wavelength gravity field, whereas the other two free parameters can be determined iteratively by the use of two additional constraints. In practice, the mantle density and mean crustal thickness are varied in order to match the seismically determined crustal thickness and to ensure that the minimum crustal thickness is equal to a chosen value. The most recent interpretations of the Apollo seismic data imply a crustal thickness of either 30 ± 2.5 km (Lognonné et al., 2003) or 38 ± 8 (Khan and Mosegaard, 2002) in the vicinity of the Apollo 12 and 14 landing sites (see also Chenet et al., 2006). When using the constraint that the minimum crustal thickness should be close to zero, as implied by the possible excavation of mantle materials from some impact basins (Yamamoto et al., 2010), the seismic analyses imply an average crustal thickness of either 34 or 43 km (Wieczorek et al., 2013). These average thicknesses imply that the bulk abundance of alumina in the Moon is similar to that of the bulk silicate Earth.

The most notable feature of the global crustal thickness models is the dramatic thinning of the crust beneath the large impact basins. This is a natural consequence of the large quantity of material that is excavated ballistically during the impact process (e.g., Wieczorek and Phillips, 1999), and it is seen that the depth of excavation reaches several tens of kilometers beneath the largest basins. The crustal thickness is close to zero beneath both the Crisium basin on the nearside hemisphere and the Moscoviense basin on the farside hemisphere (see also Ishihara et al., 2009). When combined with the detections of dunite exposures surrounding these two basins, it is likely that these impacts excavated through the entire crust and into the underlying mantle. Despite the great size of the giant South Pole–Aitken basin on the farside, the crust is predicted to be about 20 km thick in its interior. Either the depth of excavation during this basin-forming event was atypically shallow or an impact-generated melt pool differentiated to form a low density crust (Vaughan et al., 2013). The ~1.9 km offset of the Moon's center of mass from its center of figure,

combined with the lateral variations in crustal density as implied by remote sensing data, implies that the farside crust is thicker than the nearside hemisphere by about 8 km.

The majority of the lunar impact basins are characterized by having low elevations, a large central positive free-air gravitational anomaly, a gravity low surrounding the central high, and a gravity high that outlines the basin's topographic rim. Historically, basins with central gravity highs have been referred to as *mascon* basins. The central positive gravitational anomalies are a combination of both uplift of the underlying crust–mantle interface and flexural support of surface mare basalt flows. Based on estimates of the mare basalt thicknesses, which can reach a few kilometers within the central portions of some impact basins, the crust–mantle interface has been shown to be uplifted above its pre-mare isostatic position (e.g., Hikida and Wieczorek, 2007; Neumann et al., 1996; Wieczorek and Phillips, 1999). This hypothesis is supported by the existence of mascon basins that lack evidence of mare volcanism (Konopliv et al., 1998; Zuber et al., 2013b). The surrounding annulus of negative gravity implies that this region of the lithosphere is subisostatic (Andrews-Hanna, 2013), and hydrocode simulations of the basin formation process show that this region of the lithosphere was pushed below its isostatic level during basin formation (Melosh et al., 2013). Subsequent uplift of this subisostatic lithosphere by viscous flow is responsible for uplifting the central portion of the basin above its isostatic level (Andrews-Hanna, 2013; Melosh et al., 2013).

In contrast to the largest impact basins, intermediate-sized craters have negative gravitational anomalies and generally show some form of compensation. Only about 15% of the craters in the Reindler and Arkani-Hamed (2001) study appear to be completely uncompensated or to possess excess negative gravitational anomalies due to crustal brecciation. For these intermediate-sized craters, there does not appear to be any correlation of compensation state with crater age or location.

A few studies have attempted to place constrains on the elastic thickness of the Moon using both gravity and topography data (e.g., Arkani-Hamed, 1998; Crosby and McKenzie, 2005; Sugano and Heki, 2004). Most analyses have concentrated on the mascon basins, but it is often difficult to quantify the validity of the assumptions employed in the loading models. Recent models of mascon formation invoke a two-stage process (basin formation, followed by isostatic adjustment), and it is not trivial to incorporate these aspects into standard analytic flexure models. Furthermore, the thickness of the mare fill is oftentimes not known, and this is critical for fitting the present-day central gravitational anomaly. An alternative method for estimating the elastic thickness is by comparing the location of tectonic features (such as faults and graben) to that predicted by a specified loading model (e.g., Solomon and Head, 1980). Detailed modeling of the Serenitatis basin (Freed et al., 2001) suggests that its elastic thickness was about 25 km when the concentric rilles formed and probably greater than 70 km when the younger compressional ridges formed.

Modeling of the lunar gravity field shows that there are several features related to subsurface density anomalies. As one example, two positive Bouguer gravitational anomalies are found in the vicinity of the Marius Hills volcanic complex (Kiefer, 2013). These positive anomalies most likely represent basaltic intrusions that are more dense than the surrounding crust. If these are modeled as sills, they need to be several kilometers thick and could potentially represent the magma chamber that is responsible for the numerous volcanic domes found in this region. Another example of subsurface density anomalies is linear gravitational anomalies, many of which are several hundred kilometers in length (Andrews-Hanna et al., 2013). These linear anomalies are most visible when plotting horizontal gradients of the gravity tensor (cf. Petrovskaya and Vershkov, 2006) and are plausibly the expression of ancient magmatic dikes that formed when the lunar lithosphere was in a state of extension.

Finally, one curious large-scale feature of Moon is the amplitude of its degree-2 gravity and topography terms. If the Moon were in hydrostatic equilibrium, the amplitude of the C_{20} and C_{22} terms would be directly relatable to the Earth–Moon separation (cf. eqn [21]). The present-day magnitudes of these coefficients, however, are much greater than would be expected for equilibrium at the present time. This has led to the suggestion that the equilibrium shape of the Moon was frozen into the lithosphere when it was closer to Earth early in its orbital evolution (e.g., Jeffreys, 1976; Lambeck and Pullan, 1980). Alternatively, it is possible that this shape is simply a result of large-scale crustal thickness variations or lateral variations in mantle density.

If the observed shape is interpreted as a relict equilibrium shape, then this can be achieved for an Earth–Moon separation of 22 Earth radii with an orbital eccentricity of 0.49 (the current separation and eccentricity are about 60 Earth radii and 0.0549, respectively). If the Moon was in a 3/2 spin–orbit resonance (like Mercury), then alternative solutions exist for an Earth–Moon separation of about 25 Earth radii with eccentricities as low as 0.17. These interpretations are somewhat problematic as the lunar orbit is predicted to have receded beyond this distance in less than ~100 My after the formation of the Earth–Moon system (e.g., Webb, 1982). Furthermore, Ćuk (2011) pointed out that capture into either the synchronous or 3/2 resonance is difficult for the proposed high eccentricities.

10.05.9 Future Developments and Concluding Remarks

The gravitational fields and topography of the terrestrial planets have become increasingly better characterized since the discovery of lunar 'mascons' by Muller and Sjogren in 1968. Whereas the early datasets were quite sparse, the gradual accumulation of data with each successive space mission has given rise to near-global gravity and topography spherical harmonic models. Some of the gravity models now possess spherical harmonic bandwidths greater than 900, and future missions will certainly lead to continued improvements.

Not only has the resolution of the planetary datasets continued to improve with time, but also so have the analysis techniques. Early investigations were often restricted to analyses of individual 1-D line-of-sight gravitational acceleration profiles. As data coverage became more dense, 2-D regional models were developed that were more often than not

analyzed using Cartesian techniques developed for Earth. Because of the small size of some planetary bodies, the assumption of Cartesian geometry is not always justified, and spherical analysis methods have proven to be superior. The full suite of Cartesian gravity–topography analysis techniques has been developed for the sphere, including multitaper spectral analysis, wavelet analysis, the rapid calculation of gravitational anomalies from finite-amplitude topographic relief, and realistic admittance models that take into account surface and subsurface loading. Though the approximation of Cartesian geometry may not incur large errors for some small-scale investigations, it is just as easy to use a spherical-based method that possesses a comparable computational speed.

Much has been learned about the crustal and lithospheric structures of the terrestrial planets, but there is still much to be done. In particular, in hindsight, it is now clear that many gravity–topography admittance and coherence studies have used analysis techniques whose robustness are sometimes difficult to assess. Shortcomings include incorrect application of the multitaper spectral analysis technique, the neglect of either the admittance or correlation function, and the use of a theoretical admittance model that might be an oversimplification of reality. Few studies, even for Earth, have performed these analyses entirely correctly, and one should assess critically each of the elastic thickness estimates that have been published for regions where subsurface loading is important.

It is also important to note that the concept of an elastic lithosphere, which is often used in gravitational analyses, is in actuality a gross oversimplification of reality. Because of the ease of generating a time-invariant flexural profile from a load emplaced on an elastic plate, one would like to hope that the obtained 'effective elastic thickness' has some physical meaning. Though this might be the case for regions where the magnitude of surface and subsurface loading is small and where in-plane forces are absent, it has been demonstrated that the use of a more realistic rheology can yield flexural profiles that sometimes are quite different. Unfortunately, the most realistic EVP models are computationally expensive and are not currently amenable to a robust inversion procedure using gravity and topography as constraints. A simpler elastic–plastic formulation could be used in such an inversion, but this rheological model utilizes assumptions that still might turn out to be too simplistic. Regardless, it would be appropriate to develop an elastic–plastic loading model similar to the elastic model described in **Section 10.05.7** for the sphere. One benefit of such a model is that it would be possible to invert for the regional heat flow. An additional avenue of future research is to compare the locations of surface faulting with those predicted from elastic, elastic–plastic, and EVP models.

Finally, it is worth mentioning that significant improvements to our knowledge of the gravitational fields and topography of the terrestrial planets can be expected in the future. In particular, even though the land-based topography for Earth is now known to high accuracy, there are still gaps near the polar regions that could be filled by data obtained by the orbiting Geoscience Laser Altimeter System (GLAS) laser altimeter (e.g., Abshire et al., 2005; Schutz et al., 2005; Shuman et al., 2006). Data obtained from the terrestrial mission GOCE, which contains a gravity gradiometer, are leading to improved models of the terrestrial gravity field. An area of active research for Earth, but also for the other planets in a more limited sense, is that of measuring and modeling time-variable gravity signatures that are a result of hydrologic processes and tidal deformation.

With respect to the other planets, the topography and gravity field of Mercury are continuing to be characterized by data collected by the MESSENGER spacecraft (Solomon et al., 2001), and these will be supplemented by data that will be collected by the future BepiColombo mission (Spohn et al., 2001). The JUpiter ICy moons Explorer (JUICE) will collect altimetric and gravity data for the icy satellite Ganymede and to a limited extent Europa and Ganymede. Limited gravitational data will also be collected during the end of the Venus Express mission and during the MAVEN mission to Mars.

Acknowledgments

I am indebted to Greg Neumann and Frederik Simons for their numerous discussions relating to the more subtle details of gravity and topography acquisition and analysis. Francis Nimmo, Bruce Bills, Alex Konopliv, David Sandwell, Erwan Mazarico, and Martin Pauer are kindly thanked for their constructive reviews and discussions that helped improve this manuscript. Most figures were created using the Generic Mapping Tools of Wessel and Smith (1991).

References

Abe S, Mukai T, Hirata N, et al. (2006) Mass and local topography measurements of Itokawa by Hayabusa. *Science* 312: 1344–1349. http://dx.doi.org/10.1126/science.1126272.

Abshire JB, Sun X, Riris H, et al. (2005) Geoscience Laser Altimeter System (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters* 32: L21S02. http://dx.doi.org/10.1029/2005GL024028.

Albert R and Phillips RJ (2000) Paleoflexure. *Geophysical Research Letters* 27: 2385–2388.

Albert R, Phillips RJ, Dombard A, and Brown C (2000) A test of the validity of yield strength envelopes with an elastoviscoplastic finite element model. *Geophysical Journal International* 140: 399–409.

Amante C and Eakins BW (2009) ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis. NOAA Technical Memorandum NESDIS NGDC-24, 19 pp.

Anderson FS and Smrekar SE (2006) Global mapping of crustal and lithospheric thickness on Venus. *Journal of Geophysical Research* 111: E08006. http://dx.doi.org/10.1029/2004JE002395.

Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The mass, gravity field, and ephemeris of Mercury. *Icarus* 71: 337–349.

Anderson JD, Jacobson RA, Lau EL, Moore WB, and Schubert G (2001a) Io's gravity field and interior structure. *Journal of Geophysical Research* 106: 32963–32969.

Anderson JD, Jacobson RA, McElrath TP, Moore WB, and Schubert G (2001b) Shape, mean radius, gravity field, and interior structure of Callisto. *Icarus* 153: 157–161.

Anderson JD, Jurgens RF, Lau EL, Slade MA III, and Schubert G (1996a) Shape and orientation of Mercury from radar ranging data. *Icarus* 124: 690–697.

Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996b) Gravitational constraints on the internal structure of Ganymede. *Nature* 384: 541–543.

Anderson JD and Schubert G (2010) Rhea's gravitational field and interior structure inferred from archival data files of the 2005 Cassini flyby. *Physics of the Earth and Planetary Interiors* 178: 176–182. http://dx.doi.org/10.1016/j.pepi.2009.09.003.

Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998) Europa's differentiated internal structure: Inferences from four Galileo encounters. *Science* 281: 2019–2022.

André SL, Watters TR, and Robinson MS (2005) The long wavelength topography of Beethoven and Tolstoj basins, Mercury. *Geophysical Research Letters* 32: L21202. http://dx.doi.org/10.1029/2005GL023627.

Andrews-Hanna JC (2013) The origin of the non-mare mascon gravity anomalies in lunar basins. *Icarus* 222: 159–168. http://dx.doi.org/10.1016/j.icarus.2012.10.031.

Andrews-Hanna JC, Asmar SW, Head JW, et al. (2013) Ancient igneous intrusions and early expansion of the Moon revealed by GRAIL gravity gradiometry. *Science* 339: 675–678. http://dx.doi.org/10.1126/science.1231753.

Araki H, Tazawa S, Nodo H, et al. (2009) Lunar global shape and polar topography derived from Kaguya-LALT laser altimetry. *Science* 323: 897–900. http://dx.doi.org/10.1126/science.1164146.

Archinal BA, A'Hearn MF, Bowell E, et al. (2011a) Report of the IAU working group on cartographic coordinates and rotational elements: 2009. *Celestial Mechanics and Dynamical Astronomy* 109: 101–135. http://dx.doi.org/10.1007/s10569-010-9320-4.

Archinal BA, A'Hearn MF, Bowell E, et al. (2011b) Erratum to: Reports of the IAU working group on cartographic coordinates and rotational elements: 2006 & 2009. *Celestial Mechanics and Dynamical Astronomy* 110: 401–403. http://dx.doi.org/10.1007/s10569-011-9362-2.

Ardalan AA, Karimi R, and Grafarend EW (2010) A new reference equipotential surface, and reference ellipsoid for the planet Mars. *Earth, Moon, and Planets* 106: 1–13.

Arkani-Hamed J (1998) The lunar mascons revisited. *Journal of Geophysical Research* 103: 3709–3739.

Audet P (2011) Directional wavelet analysis on the sphere: Application to gravity and topography of the terrestrial planets. *Journal of Geophysical Research* 116: E01003. http://dx.doi.org/10.1029/2010JE003710.

Audet P (2013) Toward mapping the effective elastic thickness of planetary lithospheres from a spherical wavelet analysis of gravity and topography. *Physics of the Earth and Planetary Interiors* 226: 48–82. http://dx.doi.org/10.1016/j.pepi.2013.09.011.

Balmino G (1993) The spectra of the topography of the Earth, Venus, and Mars. *Geophysical Research Letters* 20: 1063–1066.

Balmino G (1994) Gravitational potential harmonics from the shape of an homogeneous body. *Celestial Mechanics and Dynamical Astronomy* 60: 331–364.

Balmino G, Vales N, Bonvalot S, and Briais A (2012) Spherical harmonic modelling to ultra-high degree of Bouguer and isostatic anomalies. *Journal of Geodesy* 86: 499–520. http://dx.doi.org/10.1007/s00190-011-0533-4.

Banerdt WB (1986) Support of long-wavelength loads on Venus and implications for internal structure. *Journal of Geophysical Research* 91: 403–419.

Banerdt WB, Golombek M, and Tanaka K (1992) Stress and tectonics on Mars. In: Jakosky B, Snyder C, Matthews M, and Kieffer H (eds.) *Mars*, pp. 249–297. Tucson, AZ: University of Arizona Press.

Banks RJ, Francis SC, and Hipkin RG (2001) Effects of loads in the upper crust on estimates of the elastic thickness of the lithosphere. *Geophysical Journal International* 145: 291–299.

Barnett DN, Nimmo F, and McKenzie D (2000) Elastic thickness estimates for Venus using line of sight accelerations from Magellan cycle 5. *Icarus* 146: 404–419.

Barnett DN, Nimmo F, and McKenzie D (2002) Flexure of Venusian lithosphere measured from residual topography and gravity. *Journal of Geophysical Research* 107: E2. http://dx.doi.org/10.1029/2000JE001398.

Bechtel TD, Forsyth DW, Sharpton VL, and Grieve RAF (1990) Variations in effective elastic thickness of the North-American lithosphere. *Nature* 343: 636–638.

Bechtel TD, Forsyth DW, and Swain CJ (1987) Mechanisms of isostatic compensation in the vicinity of the East African Rift, Kenya. *Geophysical Journal of the Royal Astronomical Society* 90: 445–465.

Belleguic V, Lognonné P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005. http://dx.doi.org/10.1029/2005JE002437.

Beuthe M (2008) Thin elastic shells with variable thickness for lithospheric flexure of one-plate planets. *Geophysical Journal International* 171: 817–841. http://dx.doi.org/10.1111/j.1365-246X.2007.03671.x.

Beuthe M, Maistre SL, Rosenblatt P, Pätzold M, and Dehant V (2012) Density and lithospheric thickness of the Tharsis Province from MEX MaRS and MRO gravity data. *Journal of Geophysical Research* 117: E04002. http://dx.doi.org/10.1029/2011JE003976.

Bills BG and Lemoine FG (1995) Gravitational and topographic isotropy of the Earth, Moon, Mars, and Venus. *Journal of Geophysical Research* 100: 26275–26295.

Blakely RJ (1995) *Potential Theory in Gravity and Magnetic Applications*. New York: Cambridge University Press.

Brown CD and Grimm RE (1996) Lithospheric rheology and flexure at Artemis Chasma, Venus. *Journal of Geophysical Research* 101: 12697–12708.

Brown CD and Phillips RJ (2000) Crust–mantle decoupling by flexure of continental lithosphere. *Journal of Geophysical Research* 105: 13221–13237.

Burov EB and Diament M (1995) The effective elastic thickness (T_e) of continental lithosphere: What does it really mean? *Journal of Geophysical Research* 100: 3905–3927.

Ceuleneer G, Rabinowicz M, Monnereau M, Cazenave A, and Rosemberg C (1988) Viscosity and thickness of the sub-lithospheric low-viscosity zone: Constraints from geoid and depth over oceanic swells. *Earth and Planetary Science Letters* 89: 84–102.

Chambat F and Valette B (2005) Earth gravity up to second order in topography and density. *Physics of the Earth and Planetary Interiors* 151: 89–106.

Chambodut A, Panet I, Mandea M, Diament M, Holschneider M, and Jamet O (2005) Wavelet frames: An alternative to spherical harmonic representation of potential fields. *Geophysical Journal International* 163: 875–899. http://dx.doi.org/10.1111/j.1365-246X.2005.02754.x.

Chao BF and Rubincam DP (1989) The gravitational field of Phobos. *Geophysical Research Letters* 16: 859–862.

Chenet H, Lognonné P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from the Apollo seismic data set. *Earth and Planetary Science Letters* 243: 1–14.

Cheng AF, Barnouin-Jha O, Prockter L, et al. (2002) Small-scale topography of 433 Eros from laser altimetry and imaging. *Icarus* 155: 51–74.

Clark PE, Jurgens RF, and Leake MA (1988) Goldstone radar observations of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 77–100. Tucson, AZ: University of Arizona Press.

Correia ACM and Laskar J (2009) Mercury's capture into the 3/2 spin–orbit resonance including the effect of core–mantle friction. *Icarus* 201: 1–11. http://dx.doi.org/10.1016/j.icarus.2008.12.034.

Crosby A and McKenzie D (2005) Measurements of the elastic thickness under ancient lunar terrain. *Icarus* 173: 100–107.

Čuk M (2011) Lunar shape does not record a past eccentric orbit. *Icarus* 211: 97–100. http://dx.doi.org/10.1016/j.icarus.2010.08.027.

Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.

Davies ME and Batson RM (1975) Surface coordinates and cartography of Mercury. *Journal of Geophysical Research* 80: 2417–2430.

Dermott SF (1979) Shapes and gravitational moments of satellites and asteroids. *Icarus* 37: 575–586.

Dermott SF and Thomas PC (1988) The shape and internal structure of Mimas. *Icarus* 73: 25–65. http://dx.doi.org/10.1016/0019-1035(88)90084-X.

Dorman LM and Lewis BT (1970) Experimental isostasy: 1. Theory of the determination of the Earth's isostatic response to a concentrated load. *Journal of Geophysical Research* 75: 3357–3365.

Drinkwater M, Floberghagen R, Haagmans R, Muzi D, and Popescu A (2003) GOCE: ESA's first Earth Explorer Core mission. *Space Science Reviews* 108: 419–432. http://dx.doi.org/10.1023/A:1026104216284.

Driscoll JR and Healy DM (1994) Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics* 15: 202–250. http://dx.doi.org/10.1006/aama.1994.1008.

Ebinger C, Bechtel T, Forsyth D, and Bowin C (1989) Effective elastic plate thickness beneath the east African and Afar plateaus and dynamic compensation of the uplifts. *Journal of Geophysical Research* 94: 2883–2901.

Esposito P, Benerdt W, Lindal G, Sjogren W, and Slade M (1992) Gravity and topography. In: Jakosky B, Snyder C, Matthews M, and Kieffer H (eds.) *Mars*, pp. 209–248. Tucson, AZ: University of Arizona Press.

Fedi M and Florio G (2002) A stable downward continuation by using the ISVD method. *Geophysical Journal International* 151: 146–156.

Fengler MJ, Freeden W, Kohlhas A, Michel V, and Peters T (2007) Wavelet modeling of regional and temporal variations of the Earth's gravitational potential observed by GRACE. *Journal of Geodesy* 81: 5–15. http://dx.doi.org/10.1007/s00190-006-0040-1.

Floberghagen R, Fehringer M, Lamarre D, et al. (2011) Mission design, operation and exploitation of the gravity field and steady-state ocean circulation explorer mission. *Journal of Geodesy* 85: 749–758. http://dx.doi.org/10.1007/s00190-011-0498-3.

Ford P and Pettengill G (1992) Venus topography and kilometer-scale slopes. *Journal of Geophysical Research* 97: 13103–13114.

Förste C, Bruinsma SL, Flechtner F, et al. (2012) A new release of EIGEN-6C. In: *AGU Fall Meeting, San Francisco, 3–7 December, G31B–0923*.

Forsyth DW (1985) Subsurface loading and estimates of the flexural rigidity of continental lithosphere. *Journal of Geophysical Research* 90: 12623–12632.

Freed AM, Melosh HJ, and Solomon SC (2001) Tectonics of mascon loading: Resolution of the strike-slip faulting paradox. *Journal of Geophysical Research* 106: 20603–20620.

Freeden W and Windheuser U (1997) Combined spherical harmonic and wavelet expansion—A future concept in Earth's gravitational determination. *Applied and Computational Harmonic Analysis* 4: 1–37.

Fukushima T (2011) Numerical computation of spherical harmonics of arbitrary degree and order by extending exponent of floating point numbers. *Journal of Geodesy* 86: 271–285. http://dx.doi.org/10.1007/s00190-011-0519-2.

Garmier R and Barriot JP (2001) Ellipsoidal harmonic expansions of the gravitational potential: Theory and application. *Celestial Mechanics and Dynamical Astronomy* 79: 235–275.
Garrick-Bethell I, Wisdom J, and Zuber MT (2006) Evidence for a past high-eccentricity lunar orbit. *Science* 313: 652–655. http://dx.doi.org/10.1126/science.1128237.
Gaskell RW, Barnouin-Jha OS, Scheeres DJ, et al. (2008) Characterizing and navigating small bodies with imaging data. *Meteoritics and Planetary Science* 43: 1049–1061. http://dx.doi.org/10.1111/j.1945-5100.2008.tb00692.x.
Genova A, Iess L, and Marabucci M (2013) Mercury's gravity field from the first six months of MESSENGER data. *Planetary and Space Science* 81: 55–64. http://dx.doi.org/10.1016/j.pss.2013.02.006.
Ghent ED, Dipple GM, and Russell JK (2004) Thermodynamic models for eclogite mantle lithosphere. *Journal of Geophysical Research* 218: 451–462.
Giese B, Denk T, Neukum G, et al. (2008) The topography of Iapetus' leading side. *Icarus* 193: 359–371. http://dx.doi.org/10.1016/j.icarus.2007.06.005.
Giese B, Oberst J, Roatsch T, Neukum G, Head JW, and Pappalardo R (1998) The local topography of Uruk Sulcus and Galileo Regio obtained from stereo images. *Icarus* 135: 303–316.
Goossens S, Matsumoto K, Liu Q, et al. (2011) Lunar gravity field determination using selene same-beam differential VLBI tracking data. *Journal of Goedesy* 85: 205–228. http://dx.doi.org/10.1007/s00190-010-0430-2.
Grimm RE (1994) The deep structure of Venusian plateau highlands. *Icarus* 112: 89–103.
Grimm RE and Hess PC (1997) The crust of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1205–1244. Tucson, AZ: University of Arizona Press.
Grimm RE and Phillips RJ (1991) Gravity anomalies, compensation mechanisms, and the geodynamics of Western Ishtar Terra, Venus. *Journal of Geophysical Research* 96: 8309–8324.
Grott M, Baratoux D, Hauber E, et al. (2013) Long-term evolution of the Martian crust–mantle system. *Space Science Reviews* 174: 49–111. http://dx.doi.org/10.1007/s11214-012-9948-3.
Grott M and Wieczorek MA (2012) Density and lithospheric structure at Tyrrhena Patera, Mars, from gravity and topography data. *Icarus* 221: 43–52. http://dx.doi.org/10.1016/j.icarus.2012.07.008.
Guilloux F, Faÿ G, and Cardoso JF (2009) Practical wavelet design on the sphere. *Applied and Computational Harmonic Analysis* 26: 143–160. http://dx.doi.org/10.1016/j.acha.2008.03.003.
Haines GV (1985) Spherical cap harmonic analysis. *Journal of Geophysical Research* 90: 2583–2591.
Han SC, Mazarico E, and Lemoine FG (2009) Improved nearside gravity field of the Moon by localizing the power law constraint. *Geophysical Research Letters* 36: L11203. http://dx.doi.org/10.1029/2009GL038556.
Hansen VL and Phillips RJ (1995) Formation of Ishtar Terra, Venus: Surface and gravity constraints. *Geology* 23: 292–296.
Harmon JK, Campbell DB, Bindschadler DL, Head JW, and Shapiro II (1986) Radar altimetry of Mercury: A preliminary analysis. *Journal of Geophysical Research* 91: 385–401.
Haruyama J, Matsunaga T, Ohtake M, et al. (2008) Global lunar-surface mapping experiment using the lunar imager/spectrometer on selene. *Earth Planets Space* 60: 243–255.
Hastings DA and Dunbar PK (1999) Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Documentation, Volume 1.0. Key to Geophysical Records Documentation (KGRD) 34. Boulder, CO: National Oceanic and Atmospheric Administration, National Geophysical Data Center.
Hauck SA II, Phillips RJ, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103: 13635–13642.
Haxby WF and Turcotte DL (1978) On isostatic geoid anomalies. *Journal of Geophysical Research* 83: 5473–5478.
Heiskanen WA and Moritz H (1967) *Physical Geodesy*. San Francisco, CA: Freeman.
Hikida H and Wieczorek MA (2007) Crustal thickness of the Moon: New constraints from gravity inversions using polyhedral shape models. *Icarus* 192: 150–166. http://dx.doi.org/10.1016/j.icarus.2007.06.015.
Hirt C (2012) Efficient and accurate high-degree spherical harmonic synthesis of gravity field functionals at the Earth's surface using the gradient approach. *Journal of Geodesy* 86: 729–744. http://dx.doi.org/10.1007/s00190-012-0550-y.
Hirt C, Featherstone WE, and Göttl F (2012) Topographic/isostatic evaluation of new-generation GOCE gravity field models. *Journal of Geophysical Research* 117: B05407. http://dx.doi.org/10.1029/2011JB008878.
Hofmann-Wellenhof B and Moritz H (2006) *Physical Geodesy*, 2nd edn. Wien, Austria: Springer Wien-New York.
Holmes SA and Featherstone WE (2002) A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions. *Journal of Geodesy* 76: 279–299.
Holschneider M, Chambodut A, and Mandea M (2003) From global to regional analysis of the magnetic field on the sphere using wavelet frames. *Physics of the Earth and Planetary Interiors* 135: 107–124. http://dx.doi.org/10.1016/S0031-9201(02)00210-8.
Holschneider M and Iglewska-Nowak I (2007) Poisson wavelets on the sphere. *Journal of Fourier Analysis and Applications* 13: 405–419. http://dx.doi.org/10.1007/s00041-006-6909-9.
Hoogenboom T, Houseman TG, and Martin P (2005) Elastic thickness estimates for coronae associated with chasmata on Venus. *Journal of Geophysical Research* 110: E09003. http://dx.doi.org/10.1029/2004JE002394.
Hoogenboom T and Smrekar SE (2006) Elastic thickness estimates for the northern lowlands of Mars. *Earth and Planetary Science Letters* 248: 830–839. http://dx.doi.org/10.1016/j.epsl.2006.06.035.
Hoogenboom T, Smrekar SE, Anderson FS, and Houseman G (2004) Admittance survey of type 1 coronae on Venus. *Journal of Geophysical Research* 109: E03002. http://dx.doi.org/10.1029/2003JE002171.
Huang J, Yang A, and Zhong S (2013) Constraints of the topography, gravity and volcanism on Venusian mantle dynamics and generation of plate tectonics. *Earth and Planetary Science Letters* 362: 207–214. http://dx.doi.org/10.1016/j.epsl.2012.11.051.
Huang Q and Wieczorek MA (2012) Density and porosity of the lunar crust from gravity and topography. *Journal of Geophysical Research* 117: E05003. http://dx.doi.org/10.1029/2012JE004062.
Hwang C and Chen SK (1997) Fully normalized spherical cap harmonics: Application to the analysis of sea-level data from TOPEX/POSEIDON and ERS-1. *Geophysical Journal International* 129: 450–460.
Hwang C and Parsons B (1996) An optimal procedure for determining gravity from multi-satellite altimetry. *Geophysical Journal International* 125: 705–719.
Iess L, Rappaport NJ, Jacobson RA, et al. (2010) Gravity field, shape, and moment of inertia of Titan. *Science* 327: 1367–1369. http://dx.doi.org/10.1126/science.1182583.
Ishihara Y, Goossens S, Matsumoto K, et al. (2009) Crustal thickness of the Moon: Implications for farside basin structures. *Geophysical Research Letters* 36: L19202. http://dx.doi.org/10.1029/2009GL039708.
Jacobson R (2010) The orbits and masses of the martian satellites and the libration of phobos. *Astronomical Journal* 139: 668679. http://dx.doi.org/10.1088/0004-6256/139/2/668.
James PB, Zuber MT, and Phillips RJ (2013) Crustal thickness and support of topography on Venus. *Journal of Geophysical Research: Planets* 118: 859–875. http://dx.doi.org/10.1029/2012JE004237.
Jaumann R, Williams DA, Buczkowski DL, et al. (2012) Vesta's shape and morphology. *Science* 336: 687–690. http://dx.doi.org/10.1126/science.1219122.
Jeffreys H (1976) *The Earth: Its Origin History and Physical Constitution*, 6th edn. New York: Cambridge University Press.
Jekeli C (1988) The exact transformation between ellipsoidal and spherical harmonic expansions. *Manuscripta Geodaetica* 14: 106113.
Jekeli C, Lee JK, and Kwon JH (2007) On the computation and approximation of ultra-high-degree spherical harmonic series. *Journal of Geodesy* 81: 603–615. http://dx.doi.org/10.1007/s00190-006-0123-z.
Johnson C, Solomon S, Head JW III, Phillips RJ, Smith D, and Zuber M (2000) Lithospheric loading by the northern polar cap on Mars. *Icarus* 144: 313–328.
Johnson CL and Sandwell DT (1994) Lithospheric flexure on Venus. *Geophysical Journal International* 119: 627–647.
Kaula WM (1967) Theory of statistical analysis of data distributed over a sphere. *Reviews of Geophysics* 5: 83–107.
Kaula WM (2000) *Theory of Satellite Geodesy: Applications of Satellites to Geodesy*. Mineola, NY: Dover.
Khan A and Mosegaard K (2002) An enquiry into the lunar interior—A non-linear inversion of the Apollo lunar seismic data. *Journal of Geophysical Research* 107: 5036. http://dx.doi.org/10.1029/2001JE001658.
Kiefer W (2004) Gravity evidence for an extinct magma chamber beneath Syrtis Major, Mars: A look at the magmatic plumbing system. *Earth and Planetary Science Letters* 222: 349–361.
Kiefer W (2005) Buried mass anomalies along the hemispheric dichotomy in eastern Mars: Implications for the origin and evolution of the dichotomy. *Geophysical Research Letters* 32: L22201. http://dx.doi.org/10.1029/2005GL024260.
Kiefer WS (2013) Gravity constraints on the subsurface structure of the Marius Hills: The magmatic plumbing of the largest lunar volcanic dome complex. *Journal of Geophysical Research: Planets* 118: 733–745. http://dx.doi.org/10.1029/2012JE004111.

Kiefer WS and Hager BF (1991) A mantle plume model for the equatorial highlands of Venus. *Journal of Geophysical Research* 96: 20947–20966.

Kiefer WS, Richards MA, and Hager BH (1986) A dynamic model of Venus's gravity field. *Geophysical Research Letters* 13: 14–17.

Kirby JF and Swain CJ (2006) Mapping the mechanical anisotropy of the lithosphere using a 2D wavelet coherence, and its application to Australia. *Physics of the Earth and Planetary Interiors* 158: 122–138. http://dx.doi.org/10.1016/j.pepi.2006.03.022.

Kirby JF and Swain CJ (2009) A reassessment of spectral T_e estimation in continental interiors: The case of North America. *Journal of Geophysical Research* 114: B08401. http://dx.doi.org/10.1029/2009JB006356.

Konopliv AS, Asmar SW, Folkner WM, et al. (2011) Mars high resolution gravity fields from MRO, Mars seasonal gravity, and other dynamical parameters. *Icarus* 211: 401–428. http://dx.doi.org/10.1016/j.icarus.2010.10.004.

Konopliv AS, Asmar SW, Park RS, et al. (2013a) The vesta gravity field, spin pole and rotation period, landmark positions, and ephemeris from the dawn tracking and optical data. *Icarus* http://dx.doi.org/10.1016/j.icarus.2013.09.005, in press.

Konopliv AS, Asmar SW, and Yuan DN (2001) Recent gravity models as a result of the Lunar Prospector mission. *Icarus* 150: 1–18.

Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.

Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from Lunar Prospector. *Science* 281: 1476–1480.

Konopliv AS, Park RS, Yuan DN, et al. (2013b) The JPL lunar gravity field to spherical harmonic degree 660 from the GRAIL primary mission. *Journal of Geophysical Research: Planets* 1415–1434. http://dx.doi.org/10.1002/jgre.20097.

Konopliv AS, Yoder CF, Standish EM, Yuan DN, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 23–50.

Kraus H (1967) *Thin Elastic Shells: An Introduction to the Theoretical Foundations and the Analysis of Their Static and Dynamic Behavior*. New York: John Wiley.

Kucinskas AB and Turcotte DT (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112: 104–116.

Lambeck K (1988) *Geophysical Geodesy*. New York: Oxford University Press.

Lambeck K and Pullan S (1980) The lunar fossil bulge hypothesis revisited. *Physics of the Earth and Planetary Interiors* 22: 29–35.

Lawrence KP and Phillips RJ (2003) Gravity/topography admittance inversion on Venus using niching genetic algorithms. *Geophysical Research Letters* 30: 1994. http://dx.doi.org/10.1029/2003GL017515.

Le Feuvre M and Wieczorek MA (2011) Nonuniform cratering of the Moon and a revised crater chronology of the inner Solar System. *Icarus* 214: 1–20. http://dx.doi.org/10.1016/j.icarus.2011.03.010.

Lemoine F, Kenyon S, Factor JK, et al. (1998) *The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96*. Greenbelt, MD: NASA Goddard Space Flight Cent.

Lemoine FG, Goossens S, Sabaka TJ, et al. (2013) High-degree gravity models from GRAIL primary mission data. *Journal of Geophysical Research: Planets* 118: 1676–1698. http://dx.doi.org/10.1002/jgre.20118.

Lemoine FG, Smith DE, Rowlands DD, et al. (2001) An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor. *Journal of Geophysical Research* 106: 23359–23376.

Li TH (1999) Multiscale representation and analysis of spherical data by spherical wavelets. *SIAM Journal on Scientific Computing* 21: 924–953.

Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model for the Moon: Implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.

Lorenz RD, Stiles BW, Aharonson O, et al. (2013) A global topographic map of Titan. *Icarus* 225: 367–377. http://dx.doi.org/10.1016/j.icarus.2013.04.002.

Lowry AR and Zhong S (2003) Surface versus internal loading of the Tharsis rise, Mars. *Journal of Geophysical Research* E9: 5099. http://dx.doi.org/10.1029/2003JE002111.

Marks KM and Smith WHF (2006) An evaluation of publicly available global bathymetry grids. *Marine Geophysical Research* 27: 19–34.

Martinec Z, Pěč K, and Burša M (1989) The Phobos gravitational field modeled on the basis of its topography. *Earth, Moon, and Planets* 45: 219–235.

Matsumoto K, Goossens S, Ishihara Y, et al. (2010) An improved lunar gravity field model from SELENE and historical tracking data: Revealing the farside gravity features. *Journal of Geophysical Research* 115: E06007. http://dx.doi.org/10.1029/2009JE003499.

McEwen JD, Hobson MP, Lasenby AN, and Mortlock DJ (2007) Fast directional continuous spherical wavelet transform algorithms. *IEEE Transactions on Signal Processing* 55: 520–529.

McGovern PJ, Solomon SC, Smith DE, et al. (2002) Localized gravity/topography admittance and correlation spectra on Mars: Implications for regional and global evolution. *Journal of Geophysical Research* 107: 5136. http://dx.doi.org/10.1029/2002JE001854.

McGovern PJ, Solomon SC, Smith DE, et al. (2004) Correction to "Localized gravity/topography admittance and correlation spectra on Mars: Implications for regional and global evolution" *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2004JE002286.

McKenzie D (1994) The relationship between topography and gravity on Earth and Venus. *Icarus* 112: 55–88.

McKenzie D (2003) Estimating T_e in the presence of internal loads. *Journal of Geophysical Research* 108: 2348. http://dx.doi.org/10.1029/2002JB001766.

McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130: 198–216.

McKenzie DP, Barnett DN, and Yuan DN (2002) The relationship between Martian gravity and topography. *Earth and Planetary Science Letters* 195: 1–16.

McNutt MK (1984) Lithospheric flexure and thermal anomalies. *Journal of Geophysical Research* 89: 11180–11194.

Melosh HJ, Freed AM, Johnson BC, et al. (2013) The origin of lunar mascon basins. *Science* 1552–1555. http://dx.doi.org/10.1126/science.1235768.

Michel V and Wolf K (2008) Numerical aspects of a spline-based multiresolution recovery of the harmonic mass density out of gravity functionals. *Geophysical Journal International* 173: 1–16. http://dx.doi.org/10.1111/j.1365-246X.2007.03700.x.

Milbury CAE, Smrekar SE, Raymond CA, and Schubert G (2009) Lithospheric structure in the eastern region of Mars' dichotomy boundary. *Planetary and Space Science* 55: 280–288. http://dx.doi.org/10.1016/j.pss.2006.03.009.

Mohit PS, Johnson CL, Barnouin-Jha O, Zuber MT, and Solomon SC (2009) Shallow basins on Mercury: Evidence of relaxation? *Earth and Planetary Science Letters* 285: 355–363. http://dx.doi.org/10.1016/j.epsl.2009.04.023.

Mohr PJ, Taylor BN, and Newell DB (2012) CODATA recommended values of the fundamental physical constants, 2010. *Reviews of Modern Physics* 84: 1527–1605. http://dx.doi.org/10.1103/RevModPhys.84.1527.

Moore WB and Schubert G (1995) Lithospheric thickness and mantle/lithosphere density contrast beneath Beta Regio, Venus. *Geophysical Research Letters* 22: 429–432.

Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regressions of highland geoid and topography. *Icarus* 128: 415–428.

Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.

Mueller S and Phillips RJ (1995) On the reliability of lithospheric constraints derived from models of outer-rise flexure. *Geophysical Journal International* 123: 887–902.

Muller PM and Sjogren WL (1968) Masons: Lunar mass concentrations. *Science* 161: 680–684.

Murray CD and Dermott SF (2008) *Solar System Dynamics*. Cambridge, UK: Cambridge University Press.

Narcowich FJ and Ward JD (1996) Nonstationary wavelets on the m-sphere for scattered data. *Applied and Computational Harmonic Analysis* 3: 324–336.

National Imagery and Mapping Agency (2000) *Department of Defense World Geodetic System 1984: Its Definition and Relationship with Local Geodetic Systems*. NIMA TR8350.2, National Imagery and Mapping Agency.

Neumann GA, Rowlands DD, Lemoine FG, Smith DE, and Zuber MT (2001) Crossover analysis of Mars Orbiter Laser Altimeter data. *Journal of Geophysical Research* 106: 23753–23768.

Neumann GA, Zuber MT, Smith DE, and Lemoine FG (1996) The lunar crust: Global structure and signature of major basins. *Journal of Geophysical Research* 101: 16841–16843.

Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2004JE002262.

Nimmo F (2002) Admittance estimates of mean crustal thickness and density at the Martian hemispheric dichotomy. *Journal of Geophysical Research* 107: 5117. http://dx.doi.org/10.1029/2000JE001488.

Nimmo F, Giese B, and Pappalardo RT (2003a) Estimates of Europa's ice shell thickness from elastically-supported topography. *Geophysical Research Letters* 30: 1233. http://dx.doi.org/10.1029/2002GL016660.

Nimmo F and McKenzie D (1998) Volcanism and tectonics on Venus. *Annual Review of Earth and Planetary Sciences* 26: 23–51.

Nimmo F, Pappalardo RT, and Giese B (2003b) On the origins of band topography, Europa. *Icarus* 166: 21–32. http://dx.doi.org/10.1016/j.icarus.2003.08.002.

Nimmo F and Stevenson DJ (2001) Estimates of Martian crustal thickness from viscous relaxation of topography. *Journal of Geophysical Research* 106: 5085–5098.

Nimmo F and Tanaka K (2005) Earth crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161. http://dx.doi.org/10.1146/annurev.earth.33.092203.122637.

Nimmo F, Thomas PC, Pappalardo RT, and Moore WB (2007) The global shape of Europa: Constraints on lateral shell thickness variations. *Icarus* 191: 183–192. http://dx.doi.org/10.1016/j.icarus.2007.04.021.

Nimmo F and Watters TR (2004) Depth of faulting on mercury: Implications for heat flux and crustal and effective elastic thickness. *Geophysical Research Letters* 31: L02701. http://dx.doi.org/10.1029/2003GL018847.

Oberst J, Elgner S, Turner FS, et al. (2011) Radius and limb topography of Mercury obtained from images acquired during the MESSENGER flybys. *Planetary and Space Science* 59: 1918–1924. http://dx.doi.org/10.1016/j.pss.2011.07.003.

Ockendon JR and Turcotte DL (1977) On the gravitational potential and field anomalies due to thin mass layers. *Geophysical Journal of the Royal Astronomical Society* 48: 479–492.

Orth CP and Solomatov VS (2011) The isostatic stagnant lid approximation and global variations in the Venusian lithospheric thickness. *Geochemistry, Geophysics, Geosystems* 12: Q07018. http://dx.doi.org/10.1029/2011GC003582.

Orth CP and Solomatov VS (2012) Constraints on the Venusian crustal thickness variations in the isostatic stagnant lid approximation. *Geochemistry, Geophysics, Geosystems* 13: Q11012. http://dx.doi.org/10.1029/2012GC004377.

Pail R, Bruinsma S, Migliaccio F, et al. (2011) First GOCE gravity field models derived by three different approaches. *Journal of Geodesy* 85: 819–843. http://dx.doi.org/10.1007/s00190-011-0467-x.

Palguta J, Anderson JD, Schubert G, and Moore WB (2006) Mass anomalies on Ganymede. *Icarus* 180: 428–441. http://dx.doi.org/10.1016/j.icarus.2005.08.020.

Panasyuk SV and Hager BH (2000) Inversion for mantle viscosity profiles constrained by dynamic topography and the geoid, and their estimated errors. *Geophysical Journal International* 143: 821–836.

Panet I, Kuroishi Y, and Holschneider M (2011) Wavelet modelling of the gravity field by domain decomposition methods: An example over Japan. *Geophysical Journal International* 184: 203–219. http://dx.doi.org/10.1111/j.1365-246X.2010.04840.x.

Parker RL (1972) The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society* 31: 447–455.

Parmetier EM and Hess PC (1992) Chemical differentiation of a convecting planetary interior: Consequences for a one plate planet such as Venus. *Geophysical Research Letters* 19: 2015–2018.

Pauer M and Breuer D (2008) Constraints on the maximum crustal density from gravity topography modeling: Applications to the southern highlands of mars. *Earth and Planetary Science Letters* 276: 253–261. http://dx.doi.org/10.1016/j.epsl.2008.09.014.

Pauer M, Fleming K, and Čadek O (2006) Modeling the dynamic component of the geoid and topography of Venus. *Journal of Geophysical Research* 111: 11012. http://dx.doi.org/10.1029/2005JE002511.

Pavlis NK, Holmes SA, Kenyon SC, and Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research* 117: B04406. http://dx.doi.org/10.1029/2011JB008916.

Peale SJ (1969) Generalized Cassini's laws. *Astronomical Journal* 74: 483–489. http://dx.doi.org/10.1086/110825.

Percival DB and Walden AT (1993) *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*. New York: Cambridge University Press.

Pérez-Gussinyé M, Lowry A, Watts AB, and Velicogna I (2004) On the recovery of effective elastic thickness using spectral methods: Examples from synthetic data and from the Fennoscandian Shield. *Journal of Geophysical Research* 109: B10409. http://dx.doi.org/10.1029/2003JB002788.

Pérez-Gussinyé M, Swain CJ, Kirby JF, and Lowry AR (2009) Spatial variations of the effective elastic thickness, T_e, using multitaper spectral estimation and wavelet methods: Examples from synthetic data and application to South America. *Geochemistry, Geophysics, Geosystems* 10. http://dx.doi.org/10.1029/2008GC002229.

Perry ME, Kahan DS, Barnouin OS, et al. (2011) Measurement of the radius of Mercury by radio occultation during the MESSENGER flybys. *Planetary and Space Science* 59: 1925–1931. http://dx.doi.org/10.1016/j.pss.2011.07.022.

Petrovskaya MS and Vershkov AN (2006) Non-singular expressions for the gravity gradients in the local north-oriented and orbital reference frames. *Journal of Geodesy* 80: 117–127. http://dx.doi.org/10.1007/s00190-006-0031-2.

Phillips RJ (1994) Estimating lithospheric properties at Atla Regio. *Icarus* 112: 147–170.

Phillips RJ, Johnson CL, Mackwell SJ, Morgan P, Sandwell DT, and Zuber MT (1997) Lithospheric mechanics and dynamics of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1163–1204. Tucson, AZ: University of Arizona Press.

Phillips RJ and Lambeck K (1980) Gravity fields of the terrestrial planets: Long wavelength anomalies and tectonics. *Reviews of Geophysics* 18: 27–76.

Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.

Phipps Morgan J and Blackman DK (1993) Inversion of combined gravity and bathymetry data for crustal structure: A prescription for downward continuation. *Earth and Planetary Science Letters* 119: 167–179.

Ping J, Huang Q, Yan J, Cao J, Tang G, and Shu R (2009) Lunar topographic model CLTM-s01 from Chang'E-1 laser altimeter. *Science China Physics, Mechanics & Astronomy* 52: 1105–1114. http://dx.doi.org/10.1007/s11433-009-0144-8.

Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992) *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edn. New York: Cambridge University Press.

Preusker F, Oberst J, Head JW, et al. (2011) Stereo topographic models of Mercury after three MESSENGER flybys. *Planetary and Space Science* 59: 1910–1917. http://dx.doi.org/10.1016/j.pss.2011.07.005.

Rabus PB, Eineder M, Roth A, and Bamler R (2003) The shuttle radar topography mission—A new class of digital elevation models acquired by spaceborne radar. *Photogrammetric Engineering & Remote Sensing* 57: 241–262.

Rambaux N and Bois E (2004) Theory of the Mercury's spin–orbit motion and analysis of its main librations. *Astronomy and Astrophysics* 413: 381–393. http://dx.doi.org/10.1051/0004-6361:20031446.

Rapp RH (1989) The decay of the spectrum of the gravitational potential and the topography of the Earth. *Geophysical Journal International* 99: 449–455.

Rappaport NJ, Konopliv AS, and Kucinskas AB (1999) An improved 360 degree and order model of Venus topography. *Icarus* 139: 19–31.

Reigber C, Jochmann H, Wünsch J, et al. (2004) Earth gravity field and seasonable variability from CHAMP. In: Reigber C, Lühr H, Schwintzer P, and Wickert J (eds.) *Earth Observations with CHAMP*, pp. 25–30. Berlin: Springer.

Reindler L and Arkani-Hamed J (2001) The compensation state of intermediate size lunar craters. *Icarus* 153: 71–88.

Ritzer JA and Hauck SA II (2009) Lithospheric structure and tectonics at Isidis Planitia, Mars. *Icarus* 201: 528–539. http://dx.doi.org/10.1016/j.icarus.2009.01.025.

Roberts J and Zhong S (2004) Plume-induced topography and geoid anomalies and their implications for the Tharsis rise on Mars. *Journal of Geophysical Research* 109: E03009. http://dx.doi.org/10.1029/2003JE002226.

Robinson EM and Parsons B (1988) Effect of a shallow low-viscosity zone on the formation of midplate swells. *Journal of Geophysical Research* 93: 3144–3156.

Rodríguez E, Morris CS, Belz JE, et al. (2005). An assessment of the SRTM topographic products. Technical Report. Jet Propulsion Laboratory.

Rogers PG and Zuber MT (1998) Tectonic evolution of Bell Regio, Venus: Regional stress, lithospheric flexure, and edifice stresses. *Journal of Geophysical Research* 103: 16841–16853.

Rosenblatt P, Pinet P, and Touvenot E (1994) Comparative hypsometric analysis of Earth and Venus. *Geophysical Research Letters* 21: 465–468.

Rosenburg MA, Aharonson O, Head JW, et al. (2011) Global surface slopes and roughness of the moon from the lunar orbiter laser altimeter. *Journal of Geophysical Research* 116: E02001. http://dx.doi.org/10.1029/2010JE003716.

Rummel R (2005) Gravity and topography of Moon and planets. *Earth, Moon, and Planets* 94: 103–111. http://dx.doi.org/10.1007/s11038-005-3245-z.

Rummel R, Yi W, and Stummer G (2011) GOCE gravitational gradiometry. *Journal of Geodesy* 85: 777–790. http://dx.doi.org/10.1007/s00190-011-0500-0.

Sandwell DT and Schubert G (1992) Flexural ridges, trenches, and outer rises around coronae on Venus. *Journal of Geophysical Research* 97: 16069–16083.

Sandwell DT and Smith WHF (1997) Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *Journal of Geophysical Research* 102: 10039–10054.

Saunders RS, Pettengill GH, Arvidson RE, Sjogren WL, Johnson WTK, and Pieri L (1990) The Magellan Venus radar mapping mission. *Journal of Geophysical Research* 95: 8339–8355. http://dx.doi.org/10.1029/JB095iB06p08339.

Scheeres DJ, Khushalani B, and Werner RA (2000) Estimating asteroid density distributions from shape and gravity information. *Planetary and Space Science* 48: 965–971.

Schmidt M, Fengler M, Mayer-Gürr T, et al. (2007) Regional gravity modeling in terms of spherical base functions. *Journal of Geodesy* 81: 17–38. http://dx.doi.org/10.1007/s00190-006-0101-5.

Scholten F, Oberst J, Matz KD, et al. (2012) GId100: The near-global lunar 100 m raster DTM from LROC WAC stereo image data. *Journal of Geophysical Research* 117: E00H17. http://dx.doi.org/10.1029/2011JE003926.

Schubert G and Sandwell DT (1995) A global survey of possible subduction sites on Venus. *Icarus* 117: 173–196.

Schubert G, Solomatov VS, Tackley PJ, and Turcotte DL (1997) Mantle convection and the thermal evolution of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1245–1287. Tucson, AZ: University of Arizona Press.

Schutz BE, Zwally HJ, Shuman CA, Hancock D, and DiMarzio JP (2005) Overview of the ICESat mission. *Geophysical Research Letters* 32: L21S01. http://dx.doi.org/10.1029/2005GL024009.

Shuman C, Zwally HJ, Schutz BE, et al. (2006) ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. *Geophysical Research Letters* 33: L07501. http://dx.doi.org/10.1029/2005GL025227.

Simons FJ and Dahlen FA (2006) Spherical Slepian functions and the polar gap in geodesy. *Geophysical Journal International* 166: 1039–1061. http://dx.doi.org/10.1111/j.1365-246X.2006.03065.x.

Simons FJ, Dahlen FA, and Wieczorek MA (2006) Spatiospectral concentration on a sphere. *SIAM Review* 48: 504–536. http://dx.doi.org/10.1137/S0036144504445765.

Simons FJ, van der Hilst RD, and Zuber MT (2003) Spatio-spectral localization of isostatic coherence anisotropy in Australia and its relation to seismic anisotropy: Implications for lithospheric deformation. *Journal of Geophysical Research* 108: 2250. http://dx.doi.org/10.1029/2001JB000704.

Simons FJ, Zuber MT, and Korenaga J (2000) Isostatic response of the Australian lithosphere: Estimation of effective elastic thickness and anisotropy using multitaper spectral analysis. *Journal of Geophysical Research* 105: 19163–19184.

Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid/topography admittance of Venus. *Science* 264: 798–803.

Simons M, Solomon SC, and Hager BH (1997) Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of Venus. *Geophysical Journal International* 131: 24–44.

Sjogren WL, Banerdt WB, Chodas PW, et al. (1997) The Venus gravity field and other geodetic parameters. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1125–1161. Tucson, AZ: University of Arizona Press.

Slepian D (1983) Some comments on Fourier-analysis, uncertainty and modeling. *SIAM Review* 25: 379–393.

Smith DE, Zuber MT, Frey HV, et al. (2001a) Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research* 106: 23689–23722.

Smith DE, Zuber MT, Jackson GB, et al. (2010a) The Lunar Orbiter Laser Altimeter investigation on the Lunar Reconnaissance Orbiter mission. *Space Science Reviews* 150: 209–241. http://dx.doi.org/10.1007/s11214-009-9512-y.

Smith DE, Zuber MT, and Neumann GA (2001b) Seasonal variations of snow depth on Mars. *Science* 294: 2141–2146.

Smith DE, Zuber MT, Neumann GA, and Lemoine FG (1997) Topography of the Moon from Clementine lidar. *Journal of Geophysical Research* 102: 1591–1611.

Smith DE, Zuber MT, Neumann GA, et al. (2010b) Initial observations from 81 the Lunar Orbiter Laser Altimeter (LOLA). *Geophysical Research Letters* 37: L18204. http://dx.doi.org/10.1029/2010GL043751.

Smith DE, Zuber MT, Phillips RJ, et al. (2012) Gravity field and internal structure of Mercury from MESSENGER. *Science* 336: 214–217. http://dx.doi.org/10.1126/science.1218809.

Smith DE, Zuber MT, Phillips RJ, et al. (2010c) The equatorial shape and gravity field of Mercury from MESSENGER flybys 1 and 2. *Icarus* 209: 88–100. http://dx.doi.org/10.1016/j.icarus.2010.04.007.

Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.

Smith WHF and Sandwell DT (1997) Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277: 1956–1962.

Smith WHF and Sandwell DW (1994) Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research* 99: 21803–21824.

Smrekar SE (1994) Evidence for active hotspots on Venus from analysis of Magellan gravity data. *Icarus* 112: 2–26.

Smrekar SE, Comstock R, and Anderson FS (2003) A gravity survey of type 2 coronae on Venus. *Journal of Geophysical Research* 108: 5090. http://dx.doi.org/10.1029/2002JE001935.

Smrekar SE and Stofan ER (1999) Origin of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.

Sneeuw N (1994) Global spherical harmonic-analysis by least-squares and numerical quadrature methods in historical perspective. *Geophysical Journal International* 118: 707–716.

Solomon SC and Head JW (1980) Lunar mascon basins: Lava filling, tectonics, and evolution of the lithosphere. *Reviews of Geophysics and Space Physics* 18: 107–141.

Solomon SC and Head JW (1990) Lithospheric flexure beneath the Frejya Montes foredeep, Venus: Constraints on lithospheric thermal gradient and heat flow. *Geophysical Research Letters* 17: 1393–1396.

Solomon SC, McNutt RL, Gold RE, et al. (2001) The MESSENGER mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465. http://dx.doi.org/10.1016/S0032-0633(01)00085-X.

Spohn T, Sohl F, Wieczerkowski K, and Conzelmann V (2001) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary and Space Science* 49: 1561–1570. http://dx.doi.org/10.1016/S0032-0633(01)00093-9.

Sugano T and Heki K (2004) Isostasy of the Moon from high-resolution gravity and topography data: Implication for its thermal history. *Geophysical Research Letters* 31: L24703. http://dx.doi.org/10.1029/2004GL022059.

Tapley B, Ries J, Bettadpur S, et al. (2005) GGM02—An improved Earth gravity field model from GRACE. *Journal of Geodesy* 79. http://dx.doi.org/10.1007/s00190-005-0480-z.

Tapley BD, Bettadpur S, Ries JC, Thompson PF, and Watkins M (2004) GRACE measurements of mass variability in the Earth system. *Science* 305: 503–505.

Taylor GJ, Boynton W, Brückner J, et al. (2006) Bulk composition and early differentiation of Mars. *Journal of Geophysical Research* 111: 3. http://dx.doi.org/10.1029/2005JE002645.

Thébault E, Schott JJ, and Mandea M (2006) Revised spherical cap harmonic analysis (R-SCHA): Validation and properties. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2005JB003836.

Thébault E, Schott JJ, Mandea M, and Hoffbeck JP (2004) A new proposal for spherical cap harmonic modelling. *Geophysical Journal International* 159: 83–103.

Thomas PC (2010) Sizes, shapes, and derived properties of the Saturnian satellites after the Cassini nominal mission. *Icarus* 208: 395–401. http://dx.doi.org/10.1016/j.icarus.2010.01.025.

Thomas PC, Burns JA, Helfenstein P, et al. (2007) Shapes of the Saturnian icy satellites and their significance. *Icarus* 190: 573–584.

Thomas PC, Davies ME, Colvin TR, et al. (1998) The shape of Io from Galileo limb measurements. *Icarus* 135: 175–180. http://dx.doi.org/10.1006/icar.1998.5987.

Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70: 1055–1096.

Tsoulis D, Jamet O, Verdun J, and Gonindard N (2009) Recursive algorithms for the computation of the potential harmonic coefficients of a constant density polyhedron. *Journal of Geodesy* 83: 925–942. http://dx.doi.org/10.1007/s00190-009-0310-9.

Turcotte DL (1995) How does Venus lose heat? *Journal of Geophysical Research* 100: 16931–16940.

Turcotte DL, Willemann RJ, Haxby WF, and Norberry J (1981) Role of membrane stresses in the support of planetary topography. *Journal of Geophysical Research* 86: 3951–3959. http://dx.doi.org/10.1029/JB086iB05p03951.

U. S. Geological Survey (2002) Color-coded topography and shaded relief map of the lunar near side and far side hemispheres I-2769.

Varshalovich DA, Moskalev AN, and Khersonskii VK (1988) *Quantum Theory of Angular Momentum*. Singapore: World Scientific.

Vaughan WM, Head JW, Wilson L, and Hess PC (2013) Geology and petrology of enormous volumes of impact melt on the Moon: A case study of the Orientale basin impact melt sea. *Icarus* 223: 749–765. http://dx.doi.org/10.1016/j.icarus.2013.01.017.

Vezolainen AV, Solomatov VS, Basilevsky AT, and Head JW (2004) Uplift of Beta Regio: Three-dimensional models. *Journal of Geophysical Research* 109: E08007. http://dx.doi.org/10.1029/2004JE002259.

Vondrak R, Keller J, Chin G, and Garvin J (2010) Lunar reconnaissance orbiter (LRO): Observations for lunar exploration and science. *Space Science Reviews* 150: 7–22. http://dx.doi.org/10.1007/s11214-010-9631-5.

Wählisch M, Willner K, Oberst J, et al. (2010) A new topographic image atlas of Phobos. *Earth and Planetary Science Letters* 294: 547–553. http://dx.doi.org/10.1016/j.epsl.2009.11.003.

Wandelt BD and Górski KM (2001) Fast convolution on the sphere. *Physical Review D* 63: 123002. http://dx.doi.org/10.1103/PhysRevD.63.123002.

Ward WR (1975) Past orientation of the lunar spin axis. *Science* 189: 377–379.

Watters T (2003) Lithospheric flexure and the origin of the dichotomy boundary on Mars. *Geology* 31: 271–274.

Watts AB (2001) *Isostasy and Flexure of the Lithosphere*. Cambridge, UK: Cambridge University Press.

Watts AB and Burov EB (2003) Lithospheric strength and its relationship to the elastic and seismogenic layer thickness. *Earth and Planetary Science Letters* 213: 113–131.

Watts AB and Zhong S (2000) Observations of flexure and the rheology of oceanic lithosphere. *Geophysical Journal International* 142: 855–875.

Webb DJ (1982) Tides and the evolution of the Earth–Moon system. *Geophysical Journal of the Royal Astronomical Society* 70: 261–271.

Werner RA (1997) Spherical harmonic coefficients for the potential of a constant density polyhedron. *Computers & Geosciences* 23: 1071–1077.

Werner RA and Scheeres DJ (1997) Exterior gravitation of a polyhedron derived and compared with harmonic and mascon gravitation representations of asteroid 4769 Castalia. *Celestial Mechanics and Dynamical Astronomy* 65: 313–344.

Wessel P and Smith WHF (1991) Free software helps map and display data. *Eos, Transactions, American Geophysical Union* 72.

Wiaux Y, McEwen J, and Vielva P (2007) Complex data processing: Fast wavelet analysis on the sphere. *Journal of Fourier Analysis and Applications* 13: 477–493. http://dx.doi.org/10.1007/s00041-006-6917-9.

Wieczorek M and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.

Wieczorek MA (2008) Constraints on the composition of the Martian south polar cap from gravity and topography. *Icarus* 196: 506–517. http://dx.doi.org/10.1016/j.icarus.2007.10.026.

Wieczorek MA, Jolliff BJ, Khan A, et al. (2006) The constitution and structure of the lunar interior. In: Jolliff BJ, Wieczorek MA, Shearer CK, and Neal CR (eds.) *New Views of the Moon. Reviews in Mineralogy and Geochemistry*, vol. 60, pp. 221–364. Mineralogical Society of America.

Wieczorek MA and Le Feuvre M (2009) Did a large impact reorient the Moon? *Icarus* 200: 358–366. http://dx.doi.org/10.1016/j.icarus.2008.12.017.

Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339: 671–675. http://dx.doi.org/10.1126/science.1231530.

Wieczorek MA and Phillips RJ (1997) The structure and compensation of the lunar highland crust. *Journal of Geophysical Research* 102: 10933–10943.

Wieczorek MA and Phillips RJ (1998) Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *Journal of Geophysical Research* 103: 1715–1724. http://dx.doi.org/10.1029/97JE03136.

Wieczorek MA and Simons FJ (2005) Localized spectral analysis on the sphere. *Geophysical Journal International* 162: 655–675. http://dx.doi.org/10.1111/j.1365-246X.2005.02687.x.

Wieczorek MA and Simons FJ (2007) Minimum-variance multitaper spectral estimation on the sphere. *Journal of Fourier Analysis and Applications* 13: 665–692. http://dx.doi.org/10.1007/s00041-006-6904-1.

Wieczorek MA and Zuber MT (2004) Thickness of the Martian crust: Improved constraints from geoid-to-topography ratios. *Journal of Geophysical Research* 109: E01009. http://dx.doi.org/10.1029/2003JE002153.

Willemann RJ and Turcotte DL (1981) Support of topographic and other loads on the Moon and on the terrestrial planets. *Lunar and Planetary Science Conference* 12B: 837–851.

Williams JG, Boggs DH, and Folkner WM (2013) *DE430 Lunar Orbit, Physical Librations, and Surface Coordinates*. Jet Propulsion Laboratory, Interoffice Memorandum IOM 335-JW,DB,WF-20130722-016, July 22.

Williams JG, Konopliv AS, Boggs DH, et al. (2014) Lunar interior properties from the GRAIL mission. *Journal of Geophysical Research: Planets*, in press.

Williams JP, Ruiz J, Rosenburg MA, Aharonson O, and Phillips RJ (2011) Insolation driven variations of Mercury's lithospheric strength. *Journal of Geophysical Research* 116: E01008. http://dx.doi.org/10.1029/2010JE003655.

Yamamoto S, Nakamura R, Matsunaga T, et al. (2010) Possible mantle origin of olivine around lunar impact basins detected by SELENE. *Nature Geoscience* 3: 533–536. http://dx.doi.org/10.1038/ngeo897.

Yoder CF (1995) Astrometric and geodetic properties of Earth and the solar system. In: Ahrens TJ (ed.) *Global Earth Physics: A Handbook of Physical Constants*. *AGU Reference Shelf*, vol. 1, pp. 1–31. American Geophysical Union.

Yuan DN, Sjogren WL, Konopliv AS, and Kucinskas AB (2001) Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research* 106: 23377–23401.

Zharkov VN, Leontjev VV, and Kozenko AV (1985) Models, figures and gravitational moments of the Galilean satellites of Jupiter and icy satellites of Saturn. *Icarus* 61: 92–100.

Zhong S (2002) Effects of lithosphere on the long-wavelength gravity anomalies and their implications for the formation of the Tharsis rise on Mars. *Journal of Geophysical Research* 107: 5054. http://dx.doi.org/10.1029/2001JE001589.

Zuber M, Smith D, Lehman D, Hoffman T, Asmar S, and Watkins M (2013a) Gravity Recovery and Interior Laboratory (GRAIL): Mapping the lunar interior from crust to core. *Space Science Reviews:* 1–22. http://dx.doi.org/10.1007/s11214-012-9952-7.

Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 237–244.

Zuber MT, Bechtel TD, and Forsyth DW (1989) Effective elastic thicknesses of the lithosphere and mechanisms of isostatic compensation in Australia. *Journal of Geophysical Research* 94: 9353–9367.

Zuber MT, Phillips RJ, Andrews-Hanna JC, et al. (2007) Density of Mars' south polar layered deposits. *Science* 317: 1718–1719. http://dx.doi.org/10.1126/science.1146995.

Zuber MT and Smith DE (1997) Mars without Tharsis. *Journal of Geophysical Research* 102: 28673–28685.

Zuber MT, Smith DE, Cheng AF, et al. (2000a) The shape of 433 Eros from the NEAR-Shoemaker Laser Rangefinder. *Science* 289: 2097–2101.

Zuber MT, Smith DE, Lemoine FG, and Neumann GA (1994) The shape and internal structure of the Moon from the Clementine mission. *Science* 266: 1839–1843. http://dx.doi.org/10.1126/science.266.5192.1839.

Zuber MT, Smith DE, Phillips RJ, et al. (2012) Topography of the northern hemisphere of Mercury from MESSENGER laser altimetry. *Science* 336: 217–220. http://dx.doi.org/10.1126/science.1218805.

Zuber MT, Smith DE, Solomon SC, et al. (2008) Laser altimeter observations from MESSENGER's first Mercury flyby. *Science* 321: 77–79. http://dx.doi.org/10.1126/science.1159086.

Zuber MT, Smith DE, Watkins MM, et al. (2013b) Gravity field of the Moon from the Gravity Recovery and Interior Laboratory (GRAIL) mission. *Science* 339: 668–671. http://dx.doi.org/10.1126/science.1231507.

Zuber MT, Solomon SC, Phillips RJ, et al. (2000b) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

This page intentionally left blank

10.06 Planetary Magnetism

JEP Connerney, NASA Goddard Space Flight Center, Greenbelt, MD, USA

© 2015 Elsevier B.V. All rights reserved.

10.06.1	**Introduction**	195
10.06.2	**Tools**	196
10.06.2.1	The Offset Tilted Dipole	197
10.06.2.2	Spherical Harmonic Models	197
10.06.3	**Terrestrial Planets**	199
10.06.3.1	Mercury	199
10.06.3.1.1	Observations	199
10.06.3.1.2	Models	200
10.06.3.1.3	Discussion	202
10.06.3.2	Venus	203
10.06.3.2.1	Observations	203
10.06.3.2.2	Discussion	203
10.06.3.3	Mars	203
10.06.3.3.1	Observations	204
10.06.3.3.2	Models	207
10.06.3.3.3	Discussion	208
10.06.4	**Gas Giants**	209
10.06.4.1	Jupiter	209
10.06.4.1.1	Observations	210
10.06.4.1.2	Models	211
10.06.4.1.3	Discussion	214
10.06.4.2	Saturn	215
10.06.4.2.1	Observations	216
10.06.4.2.2	Models	216
10.06.4.2.3	Discussion	219
10.06.5	**Ice Giants**	221
10.06.5.1	Uranus	221
10.06.5.1.1	Models	222
10.06.5.2	Neptune	223
10.06.5.2.1	Models	224
10.06.6	**Satellites and Small Bodies**	225
10.06.6.1	Moon	225
10.06.6.2	Ganymede	229
10.06.6.3	Io	230
10.06.7	**Discussion**	230
10.06.8	**Summary**	231
References		232

10.06.1 Introduction

Planetary magnetism began with the study of the geomagnetic field and is one of the oldest scientific disciplines. Magnetism was both a curiosity and a practical tool for navigation, the subject of myth and superstition as much as scientific study. With the publication of *De Magnete* in 1600, William Gilbert demonstrated that the Earth's magnetic field was like that of a bar magnet. Gilbert's treatise on magnetism established the global nature and dipole geometry of the field and led to an association between magnetism and the interior of planet Earth. Observations of the Earth's magnetic field gathered over time demonstrated that it varies slowly with time. The orientation of the field was observed to execute a long-term or secular drift called polar wander. Later, when accurate measurements of the magnitude of the field could be made, secular variations in the field magnitude were identified. Given the importance of accurate magnetic field maps in navigation, global surveys of the magnetic field were conducted periodically, at first by ships of discovery and commerce and later by specialized magnetic survey ships built to minimize distortion of the magnetic field they sought to measure. In recent times, polar-orbiting, near-Earth satellites such as MAGSAT, Ørsted, CHAMP, SAC-C, and SWARM provided the capability to map the vector magnetic field over the entire surface of Earth in a brief period, to capture a snapshot of the field at each epoch. By studying the magnetic properties of terrestrial rocks, paleomagnetists can extend the history of the geodynamo over timescales of hundreds of millions of years.

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00171-8

Planetary magnetism is also a recent study, born of the space age and a new era of discovery that began in the 1960s with the launch of spacecraft to other bodies. The Sun and stars were known to possess magnetic fields, but one could only speculate on whether the planets and satellites in our solar system were magnetized. The planet Jupiter is an exception; its magnetic field was inferred from the detection of nonthermal radio emissions (Burke and Franklin, 1955) as early as 1955, before the era of planetary exploration. In the decades that followed, the gross characteristics of Jupiter's magnetic field were deduced from variations of the plane of polarization of the emissions or variations in flux density (see, e.g., the review by Berge and Gulkis (1976)). Estimates of the orientation of Jupiter's field (dipole tilt) were more successful than those of its magnitude; the detailed nature of Jupiter's field was established by direct measurement, by spacecraft.

In little more than three decades, spacecraft have visited nearly every large body in the solar system. The first *in situ* observations of the Jovian magnetosphere were obtained by the Pioneer 10 and 11 spacecraft in December 1973 and December 1974, respectively (Acuña and Ness, 1976; Smith et al., 1974, 1975a). The Pioneer 11 spacecraft continued on to make the first observations of Saturn's magnetic field as well (Acuña and Ness, 1980; Smith et al., 1980a,b). Voyagers 1 and 2 followed, passing through the Jovian magnetosphere in March and July of 1979, respectively (Ness et al., 1979a,b). Voyagers 1 and 2 continued onward to Saturn, obtaining additional observations of Saturn's magnetic field in November 1980 and August 1981 (Ness et al., 1981, 1982). Voyager 2 continued on its 'grand tour' of the solar system, with the discovery of the magnetic field of Uranus in January 1986 (Ness et al., 1986) and that of Neptune in August 1989 (Ness et al., 1989). Voyager 2's remarkable tour of the outer solar system was made possible by the celestial alignment of the planets, the syzygy, which occurs every 187 years or so.

Jupiter was visited again in February 1992 by the Ulysses spacecraft en route to the high-latitude heliosphere and again, albeit at great distance, in 2004. Ulysses used a Jupiter gravity assist to escape the ecliptic plane but in so doing passed to within $\sim 6R_j$ (Balogh et al., 1992). Additional observations were obtained by the Galileo orbiter beginning in 1995 (Johnson et al., 1992), but, apart from the deorbit maneuver to send Galileo into Jupiter's atmosphere for planetary protection (of the Galilean satellites), little additional information regarding Jupiter's main field was obtained. Instead, Galileo surprised us with the remarkable discovery of an intrinsic magnetic field of Ganymede (Kivelson et al., 1996) and the induced magnetic fields of Europa and Callisto (Khurana et al., 1998; Zimmer et al., 2000). Today, Cassini orbits Saturn gathering observations throughout the magnetosphere; observations from high-inclination orbits with periapsides near $3R_s$ are just now becoming available. Late in the extended ('Solstice') mission, one set of near-polar orbits will send Cassini through the ring plane near the F ring ($2.3R_s$), and another set ('proximal orbits') will send the spacecraft through the ring plane inside of the entire ring system, just above the cloud tops.

The Mariner 10 spacecraft discovered Mercury's magnetic field in brief flybys in 1974 and 1975 (Ness et al., 1974a, 1975, 1976) and provided the first indication that slowly rotating Venus was not magnetized (Ness et al., 1974b). As of this writing, the MESSENGER (*ME*rcury *S*urface, *S*pace *EN*vironment, *GE*ochemistry, and *R*anging) spacecraft (Gold et al., 2001; Solomon et al., 2001) continues to map the Hermean magnetic field from its vantage point, in an elliptical orbit with periapsis at high northern latitudes. The European Space Agency (ESA)'s BepiColombo mission readies two spacecraft that will orbit Mercury in the near future (Anselmi and Scoon, 2001).

A great many spacecraft visited Mars (Connerney et al., 2005), beginning with the Soviet launch of Mars 1 in 1962, but Mars did not relinquish his secrets until the Mars Global Surveyor (MGS) spacecraft entered orbit about Mars in September of 1997. MGS measurements demonstrated that Mars does not have a global magnetic field of internal origin, but its crust is intensely magnetized (Acuña et al., 1998). So Mars once had an internal magnetic field, for at least as long as it took the crust to acquire remanence. The Mars dynamo has long since disappeared, but the crust retains a magnetic imprint, acquired at least 4 billion years ago, that records at least part of the dynamo's history. The only traditional planet (setting aside the debate on the definition of 'planet') not yet visited by spacecraft is the Pluto–Charon system, which is scheduled for a visit by the New Horizons spacecraft in July 2015. The New Horizons spacecraft is not instrumented to measure magnetic fields, so no direct measurement is possible, but the presence of a magnetic field may be inferred from measurements made by other instruments.

Our knowledge of the magnetic fields of the planets is, in large part, based upon the planetary flybys that occurred decades ago. Mars is the most notable exception, having been mapped thoroughly over the course of more than 9 years by the polar-orbiting MGS spacecraft, operating at a nominal altitude of 400 km (Connerney et al., 2001, 2005). At present, orbiters are operating at both Mercury and Saturn, dramatically extending observational coverage of those planets; elsewhere, the observations are sparse in both space and time, particularly considered in light of the data available for the study of the geomagnetic field. Existing observations are, however, sufficient to describe a diverse group of planetary magnetic fields and provide clues to guide the continued development of dynamo theory. We are also likely to have a much more complete set of observations in the near future, with the Cassini Solstice Mission providing a large number of close-in, polar orbits of Saturn and BepiColombo promising nearly complete global coverage of Mercury. At this time, the Juno spacecraft is en route to Jupiter, perhaps the best target in the solar system for studies of the dynamo. Beginning with orbit insertion in July 2016, Juno will map Jupiter's magnetic field both near and far from a near-polar, elliptical orbit that ducks beneath that planet's intense radiation belts (Bolton et al., 2010).

10.06.2 Tools

A planet with an intrinsic magnetic field stands as an obstacle to the solar wind, the high velocity stream of plasma emanating from the Sun. The interaction of the planetary field and the solar wind forms a multitiered interaction region, often approximated as a set of conic sections. The supersonic solar wind forms a shock upstream of the obstacle (*bow shock*). The slowed solar wind flows around the obstacle within the *magnetosheath*, bounded by the bow shock and the *magnetopause*, often approximated by a paraboloid of revolution about the

planet–Sun line. Within the magnetopause is a region dominated by the planetary field, called the *magnetosphere*, a term introduced by Tom Gold in the postwar era of rocketry that provided access to space in the early 1950s. The magnetospheric magnetic field extends well downstream in the antisunward direction, as if stretched like an archer's bowstring, to form the magnetotail. The observed magnetic field can be regarded as the sum of contributions from several sources, dominated by the planetary dynamo, sometimes referred to as the 'main field.' Other relatively minor sources include currents flowing on the magnetopause and tail currents arising from the solar wind interaction and distributed currents (*ring currents*) due to the coherent motion of charged particles within the magnetosphere. Field-aligned currents, called *Birkeland currents*, flow between the magnetosphere and the planet's electrically conducting ionosphere, particularly during magnetospheric storms, leading to intense auroral displays. All of these currents produce magnetic fields, often relatively small when compared to the planetary field, particularly close to the planet's surface. However, if an accurate model is to be obtained from spatially limited or distant observations, these sources can be a significant source of errors if not modeled.

The planetary field is most often characterized by one of a very few simple models, depending on the application and/or the availability of observations. Simple representations include dipole models of one form or another and a potential field representation as a series of spherical harmonics. External fields, or those arising from currents outside the planet, are often accommodated by explicit models or spherical harmonic methods. The latter approach is useful only in regions free of currents.

10.06.2.1 The Offset Tilted Dipole

Dipole models have found use in the interpretation of magnetic field observations, dating to Gilbert's analysis of the Earth's magnetic field. The dipole is the simplest approximation to a localized source. The vector magnetic field $\mathbf{B}(\mathbf{r})$ of a dipole at the origin is given by

$$\mathbf{B}(\mathbf{r}) = \left[-\frac{\mathbf{m}}{r^3} + \frac{3(\mathbf{m}\cdot\mathbf{r})\mathbf{r}}{r^5}\right]$$

where $\mathbf{m}$ is the magnetic moment in units of $B - r^3$ and $r = |\mathbf{r}|$. The field of a dipole located near the origin is computed by translation of the origin of the coordinate system. Offset tilted dipole (OTD) models are found using forward modeling techniques to compare *in situ* observations to the OTD model field. The dipole offset (x_o, y_o, z_o) is not linearly related to the computed field, so an iterative procedure is often used to find the optimal dipole position (e.g., Smith et al., 1976). Alternatively, one may perform a spherical harmonic analysis (see next section) of degree and order 1 centered upon each of a large number of candidate dipole positions, selecting the offset that minimizes the model residuals. In either case, the result is the OTD that most closely approximates the field measured along the spacecraft trajectory.

The simple OTD representation provides a convenient approximation to the field, particularly useful at larger distances. It is useful in visualizing the gross characteristics of the field, and it is often used to chart the motion of charged particles throughout the magnetosphere. However, the simplicity of the OTD is also a limitation, in that complex field geometries cannot be adequately represented with such a limited parameterization. In general, a dipole representation can be expected to closely approximate the field for radial distances $r \gg a$, where a is a characteristic source dimension. At lesser radial distances, an accurate description of the field requires more flexibility in parameterization, for example, consideration of higher-degree and higher-order moments.

A logical extension of the simple OTD model is one of the many dipoles. The Earth's field has been approximated by adding one or more smaller dipoles to a 'main dipole' and treating the field as a summation over distinct sources; this provides a simple means of introducing an anomaly or departure from dipole geometry. The Earth's field has been approximated by an ensemble of many dipoles located on the surface of a sphere of radius $0.5R_e$, the fluid core radius. In addition to the field generated deep within Earth, induced and remanent magnetizations in the crust can be modeled with another set of dipoles, scattered about a sphere of radius $1.0R_e$ to approximate sources in the crust (Mayhew and Estes, 1983). A similar approach has been taken to model the crustal magnetic field of Mars (Purucker et al., 2000).

10.06.2.2 Spherical Harmonic Models

In the absence of local currents ($\nabla \times \mathbf{B} = 0$), the magnetic field may be expressed as the gradient of a scalar potential $V(\mathbf{B} = -\nabla V)$. It is particularly advantageous to expand the potential V in a series of functions, called spherical harmonics, which are solutions to Laplace's equation in spherical coordinates. This approach, introduced by Gauss in 1839, has been very popular in studies of the geomagnetic field and subsequently those of the planets. The traditional spherical harmonic expansion of V is given by (e.g., Chapman and Bartels, 1940; Langel, 1987)

$$V = a\sum_{n=1}^{\infty}\left\{\left(\frac{r}{a}\right)^n T_n^e + \left(\frac{a}{r}\right)^{n+1} T_n^i\right\}$$

where a is the planet's equatorial radius. The first series in increasing powers of $\mathbf{r}$ represents contributions due to external sources, with

$$T_n^e = \sum_{m=0}^{n}\left\{P_n^m(\cos\theta)\left[G_n^m\cos(m\phi) + H_n^m\sin(m\phi)\right]\right\}$$

The second series in inverse powers of $\mathbf{r}$ represents contributions due to the planetary field or internal sources, with

$$T_n^i = \sum_{m=0}^{n}\left\{P_n^m(\cos\theta)\left[g_n^m\cos(m\phi) + h_n^m\sin(m\phi)\right]\right\}$$

The $P_n^m(\cos\theta)$ are Schmidt quasinormalized associated Legendre functions of degree n and order m, and g_n^m, h_n^m and G_n^m, H_n^m are the internal and external Schmidt coefficients, respectively. These are most often presented in units of Gauss or nanoteslas ($1\ \mathrm{G} = 10^5\ \mathrm{nT}$) for a particular choice of equatorial radius a of the planet. Different values of the equatorial radius have been used and should be noted in the comparison of various field models. The angles θ and ϕ are the polar angles of a spherical coordinate system, θ (colatitude) measured from the axis of rotation and ϕ increasing in the direction of

rotation. The three components of the magnetic field (internal field only) are obtained from the expression for V mentioned previously:

$$B_r = -\frac{\partial V}{\partial r} = \sum_{n=1}^{\infty}\sum_{m=0}^{n}\left\{(n+1)\left(\frac{a}{r}\right)^{n+2}\left[g_n^m\cos(m\phi)+h_n^m\sin(m\phi)\right]P_n^m(\cos\theta)\right\}$$

$$B_\theta = -\frac{\partial V}{r\partial\theta} = -\sum_{n=1}^{\infty}\sum_{m=0}^{n}\left\{\left(\frac{a}{r}\right)^{n+2}\left[g_n^m\cos(m\phi)+h_n^m\sin(m\phi)\right]\frac{\mathrm{d}P_n^m(\cos\theta)}{\mathrm{d}\theta}\right\}$$

$$B_\phi = \frac{-1}{r\sin\theta}\frac{\partial V}{\partial\phi} = \frac{1}{\sin\theta}\sum_{n=1}^{\infty}\sum_{m=0}^{n}\left\{m\left(\frac{a}{r}\right)^{n+2}\left[g_n^m\sin(m\phi)-h_n^m\cos(m\phi)\right]P_n^m(\cos\theta)\right\}$$

The leading terms in the series (through degree 4) may be computed using the Schmidt quasinormalized Legendre functions listed in Table 1.

Table 1 Schmidt quasinormalized Legendre functions

$P_1^0 = (1)\cos\theta$
$P_1^1 = (1)\sin\theta$
$P_1^0 = \left(\frac{3}{2}\right)(\cos^2\theta - 1/3)$
$P_2^1 = \left(3^{1/2}\right)\cos\theta\sin\theta$
$P_2^2 = \left(\frac{3^{1/2}}{2}\right)\sin^2\theta$
$P_3^0 = \left(\frac{5}{2}\right)\cos\theta(\cos^2\theta - 9/15)$
$P_3^1 = \left(\frac{5(3)^{1/2}}{2(2)^{1/2}}\right)\sin\theta(\cos^2\theta - 3/15)$
$P_3^2 = \left(\frac{(3)^{1/2}(5)^{1/2}}{2}\right)\cos\theta\sin^2\theta$
$P_3^3 = \left(\frac{(5)^{1/2}}{2(2)^{1/2}}\right)\sin^3\theta$
$P_4^0 = \left(\frac{35}{8}\right)\left[\cos^4\theta - (30/35)\cos^2\theta + (3/35)\right]$
$P_4^1 = \left(\frac{7(5)^{1/2}}{2(2)^{1/2}}\right)\cos\theta\sin\theta(\cos^2\theta - 3/7)$
$P_4^2 = \left(\frac{7(5)^{1/2}}{4}\right)\sin^2\theta(\cos^2\theta - 1/7)$
$P_4^3 = \left(\frac{(7)^{1/2}(5)^{1/2}}{2(2)^{1/2}}\right)\cos\theta\sin^3\theta$
$P_4^4 = \left(\frac{(7)^{1/2}(5)^{1/2}}{8}\right)\sin^4\theta$

The leading term (in parentheses) converts the functions from Gaussian normalization, convenient for recursive computations, to Schmidt quasinormalized form.

Observations are often rendered in a west longitude system, in which the longitude of a stationary observer increases in time as the planet rotates. West longitudes are simply related to the angle ϕ by $\lambda = 360 - \phi$. Longitudes for the terrestrial planets may be assigned by reference to surface features. For the outer planets, the rotation period is inferred from observations of episodic radio emissions, which, by assumption, are locked in phase with the rotation of the planet's magnetic field and hence the electrically conducting deep interior of the planet. For these planets, longitudes must be assigned with knowledge of the rotation rate and time. For Jupiter, observation of high-frequency radio emission (penetrating the Earth's ionosphere) over several decades led to an accurate determination of the planet's rotation period, although one must be cognizant of the occasional update (e.g., λ_{III} (1957) vs. λ_{III} (1965)). These two systems result in significantly different longitudes when observations of the Pioneer era are compared with those of the Voyager era, so one must be careful to transform earlier results to a current coordinate system. A detailed description of Jovian coordinate systems is given by Dessler (1983). For Saturn, Uranus, and Neptune, the rotation periods and longitude system in current use are derived from Voyager observations of radio emissions (Desch and Kaiser, 1981; Desch et al., 1986; Warwick et al., 1989). In practice, the use of a radio-derived rotation period satisfies the need for a planet-fixed coordinate system for fluid planets, but exceptions (e.g., Saturn) do exist.

The expansion in increasing powers of r, representing external fields, is often truncated at $n = N_{\max} = 1$ corresponding to a uniform external field due to distant magnetopause and tail currents, that is, sources well beyond the region of interest. No potential field can represent the field of local currents, so in practice, the external expansion is often truncated at $N_{\max} = 1$ and local currents are modeled with the aid of explicit models. The dominant contribution from local currents may be described as a large-scale equatorial current system (Connerney et al., 1981). Field-aligned or Birkeland currents may also play a role in addition to induction effects but thus far have not been included in such analyses, excepting those of Earth (e.g., Backus, 1986; Sabaka et al., 2002). It is very important to adequately treat the fields of local currents, because even small errors accumulating along the spacecraft trajectory can lead to rather large errors in the spherical harmonic coefficients deduced from observations. Fields produced by external current systems vary over relatively short timescales (e.g., in response to variations in the solar wind ram pressure on the magnetopause), further complicating modeling efforts.

The maximum degree and order required of the internal field expansion depends on the complexity of the field measured. In usual practice, the series mentioned earlier is truncated at a maximum degree $N_{\max}$, where $N_{\max}$ is large enough to follow variations in the field at the orbital altitude of the measurement. The number of free parameters grows rapidly with increasing $N_{\max}$, as $n_p = (N_{\max}+1)^2 - 1$. If the observations are well distributed on a sphere, the spherical harmonics do not covary, and the coefficients obtained are independent of the truncation of the series. However, if the observations are poorly distributed or sparse, as is often the case for a planetary flyby, the usual spherical harmonics covary, so it is advisable to construct new orthogonal basis functions. Furthermore, if an

arbitrarily small choice of N_{max} is imposed upon the model in order to obtain a manageable linear system, large errors in low-degree and low-order terms will result from the neglect of higher-order terms. To address these problems, Connerney (1981) introduced a method of analysis based on the singular value decomposition (SVD) of Lanczos. The method (Lanczos, 1971) involves the construction of partial solutions to the generalized linear inverse problem to obtain estimates of those parameters that are constrained by the observations. The model parameters that are not constrained by the data are readily identified and exploited in characterization of model nonuniqueness. One advantage of this approach is that the physical model of the planetary field (expansion to degree N_{max}) does not depend on the completeness or extent of the available observations. Partial solutions, and estimates of the model parameters that result, may be interpreted within the context of the 'resolution matrix' for the particular solution (see Connerney, 1981; Connerney et al., 1991). The close flyby of Neptune in 1989 serves to illustrate this concept well.

10.06.3 Terrestrial Planets

10.06.3.1 Mercury

Mercury, the smallest ($R_m = 2439$ km) and least massive ($M_m = 0.056 M_e$) planet in the solar system, occupies the pole position in its race about the Sun, with a semimajor axis of 0.39 AU. Due to its large orbital eccentricity (0.206), in one orbit about the Sun (88 days), its orbital radius varies from 0.31 to 0.47 AU. It is locked in a spin–orbit resonance (Goldreich and Peale, 1966; Pettengill and Dyce, 1965) rotating once every 58.65 days or three times every two orbital periods. It is difficult to observe from Earth, due to its proximity to the Sun, and little was known about Mercury prior to observation by spacecraft on 29 March 1974, the first of the three encounters with the Mariner 10 spacecraft. Discovery of the intrinsic magnetic field of Mercury was a bit surprising since it was widely assumed that a body as small as Mercury would have completely solidified, ruling out generation of a magnetic field by dynamo action. The Mariner 10 observations of Mercury provided the only *in situ* observations obtained in the first four decades of space exploration. As this chapter is being written, the MESSENGER spacecraft (Gold et al., 2001; Solomon et al., 2001) is orbiting Mercury, in part to better understand the internal magnetic field and its origin. MESSENGER entered orbit in March 2011, after a series of flybys of Earth (August 2005), Venus (October 2006 and June 2007), and Mercury (three flybys: January 2008, October 2008, and September, 2009). A more ambitious mission to Mercury, 'BepiColombo,' is in development by the ESA and the Institute of Space and Astronautical Science (ISAS) in Japan for launch in the not too distant future (Anselmi and Scoon, 2001).

10.06.3.1.1 Observations

Mariner 10, known prelaunch as the Mariner Venus–Mercury spacecraft, was launched on 3 November 1973. The spacecraft used a close flyby of Venus on 5 February 1974 (for gravity assist) to achieve three close encounters with Mercury (29 March 1974, 21 September 1974, and 16 March 1975) before it exhausted its expendables. The orbital period of Mariner 10 after the Venus gravity assist was 176 days, very nearly commensurate with Mercury's orbital (and rotation) period, providing three very similar close encounters. Each time the spacecraft approached the planet from a local time of ~19.00 h and departed near 07.00 h. Given the brevity of these flyby encounters, the relatively slow planetary rotation rate, and the orbital commensurabilities, each encounter approximates a shot past a stationary object, the major difference being the altitude and subspacecraft latitude of the point of closest approach. The first and third of these encounters (M10-I and M10-III) provided observations within the Hermean magnetosphere (Figure 1); the second was targeted well above the sunlit hemisphere and beyond the reach of the Hermean magnetosphere in the upstream direction of the solar wind.

Mariner 10's first flyby was relatively close, as planetary flybys go, with a periapsis of $1.29 R_m$ over the darkened hemisphere (antisunward direction), passing from dusk to dawn behind the planet as viewed from the Sun. Mariner 10 traveled from bow shock (inbound) to bow shock (outbound) in a mere 33 min, 17 of which were spent within the confines of a Hermean magnetosphere (Figure 2). These observations are best understood in terms of the field external to the planet, even at such a close distance. The field increased in magnitude from about 45 nT just inside the magnetopause to ~100 nT, all the while maintaining the antisunward orientation characteristic of a planetary magnetotail. As Mariner 10 crossed the equator plane, the field decreased in magnitude precipitously, as if crossing a current sheet, after which it was much smaller in magnitude and chaotic, lacking the smooth variation and consistent orientation that characterized the inbound pass. These observations indicated that Mariner 10 passed through a small magnetosphere, sized just large enough to contain the solid body of the planet under nominal solar wind conditions (Ness et al., 1975). To form such a magnetosphere, Mercury must

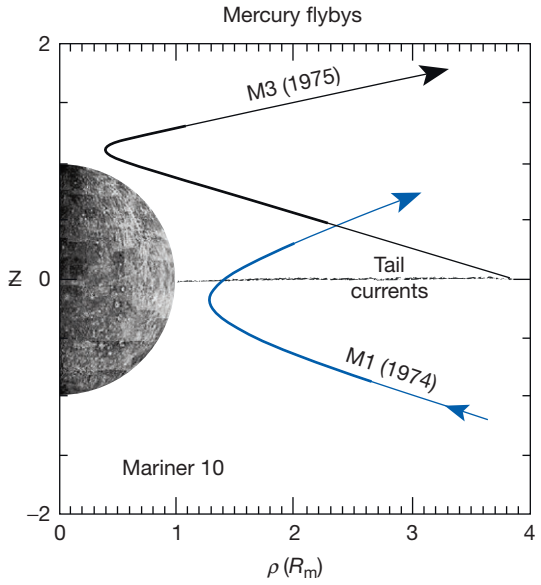


Figure 1 Mariner 10 trajectory in Mercury-centered cylindrical coordinates for the first (M1) and third (M3) encounters. The spacecraft distance from the equator (z) as a function of distance from the z-axis (ρ) is given in units of Mercury radius. The part of the trajectory within the Hermean magnetosphere is indicated by the thick line segment.

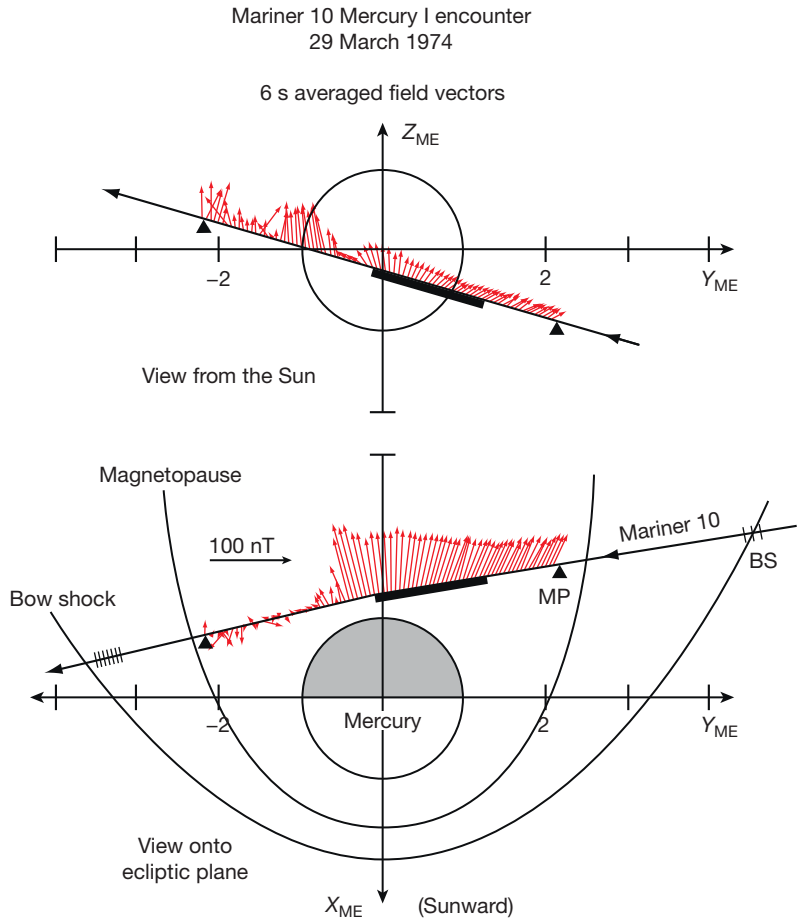


Figure 2 M1 encounter trajectory and magnetic field observations in Mercury ecliptic coordinates, viewed from the Sun (top) and viewed along Mercury's rotation axis (bottom). The vector magnetic field observed along the Mariner 10 trajectory is illustrated by vector projection onto the *z*–*y* (top) and *x*–*y* (bottom) planes; vectors originate at the spacecraft position in ME coordinates at the time of observation. Observed bow shock (BS) and magnetopause (MP) positions are shown along with scaled nominal bow shock and magnetopause boundaries.

have an internal magnetic field, albeit one with a modest dipole magnitude.

Mariner 10's second flyby was upstream of Mercury in the solar wind and too distant to penetrate the small Hermean magnetosphere. However, the third flyby was targeted to pass closely over the northern pole (Figure 3), in an effort to gather measurements optimized for characterization of the Hermean field. During the spacecraft's brief traversal of the magnetosphere (14 min), the field increased from about 20 nT to a maximum of 402 nT observed just after closest approach at $1.14R_m$ radial distance ($1R_m = 2439$ km). The approximate $1/r^3$ dependence of the observed field on spacecraft radial distance establishes the internal origin of the field, and initial analysis (Ness et al., 1976) suggested a dipole moment of ~ 342 nT-R_m^3, aligned with the $-Z$-axis (normal to the Hermean orbital plane).

The three MESSENGER flybys (M1, M2, and M3) prior to orbit insertion were quite similar to the Mariner 10's first flyby (M10-I), in that all three passed through Mercury's shadow at low latitude, sampling the internal field near the innermost extent of the cross tail current sheet (Anderson et al., 2010). The maximum total field measured during the M1 and M2 passes was ~160 nT, a magnitude that is comparable to the ~100 nT of external field produced by tail and magnetopause currents that result from the solar wind interaction (Korth et al., 2004). The third MESSENGER flyby was interrupted by a spacecraft safe mode entry near closest approach, and instruments were 'safed,' or powered off, for a brief period of time before nominal operations were restored.

MESSENGER entered orbit about Mercury on 17 March 2011 and became the first spacecraft to orbit the innermost terrestrial planet of our solar system. MESSENGER initially entered a highly elliptical, near-polar (82°–84° orbit inclination) orbit with a period of about 12 h (Figure 4); these orbits sample all local times as Mercury orbits the Sun, with closest approach altitudes ranging from 200 to 500 km at high (60°–74°) northern latitudes. After an Earth year in this orbit, the extended mission began with a pair of maneuvers (16 and 20 April 2012) that reduced the orbital period to 8 h. As of this writing, MESSENGER has completed over 2600 such orbits.

10.06.3.1.2 Models

Two fundamental issues arise in modeling the Hermean magnetic field. The first is a simple consequence of the relative

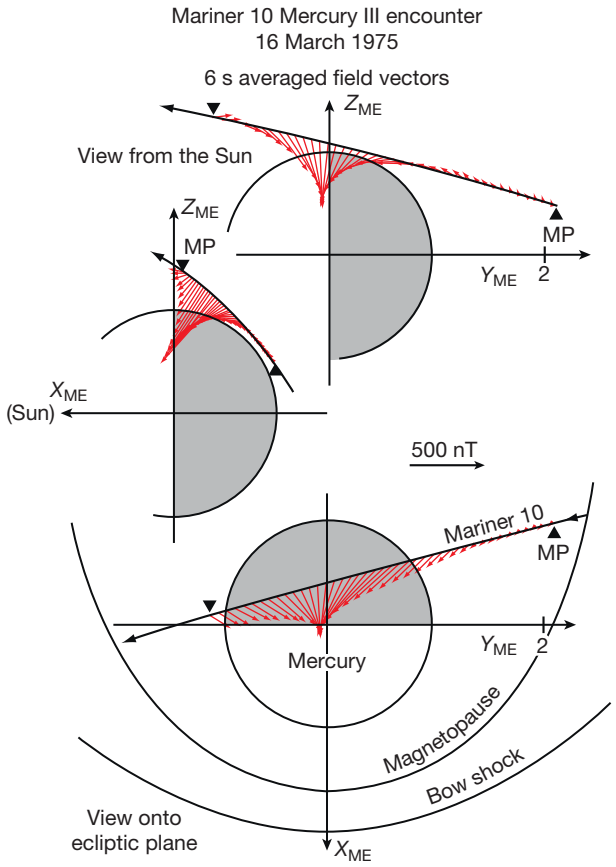


Figure 3 M3 encounter trajectory and magnetic field observations in Mercury ecliptic coordinates, viewed from the Sun (top) and viewed along Mercury's rotation axis (bottom). The vector magnetic field observed along the Mariner 10 trajectory is illustrated by vector projection onto the *z*–*y* (top), *x*–*z* (middle), and *x*–*y* (bottom) planes; vectors originate at the spacecraft position in ME coordinates at the time of observation. Observed magnetopause (MP) positions are shown along with scaled nominal bow shock and magnetopause boundaries.

weakness of the field, with a dipole moment only 0.05% that of Earth's; the second is a more subtle consequence of the distribution of observations collected by spacecraft in the immediate vicinity of the planet. With such a weak internal field, the planet occupies much of its magnetosphere, and the contribution of magnetospheric currents – primarily magnetopause currents and cross tail currents – is relatively large, time-variable, and difficult to model with great accuracy. Thus, models of Mercury's magnetic field fall largely into two groups, distinguished by their treatment of external fields. The first group of models used a spherical harmonic expansion for both internal and external fields and were limited to the dipole ($n=1$) terms for internal fields. Since these are potential field models, they are restricted to the region in which no currents flow, so they may be applied, for example, within only one lobe of the tail. Initial analyses used the first inbound Mariner 10 observations through to the tail current sheet crossing, while later analyses benefited from the Mariner 3 observations obtained crossing the pole. The latter are better suited to internal field estimation by virtue of improved signal to noise (magnitude relative to external fields). These are listed in **Table 2**.

Another set of models were obtained using a spherical harmonic expansion for the internal field, augmented by explicit models of the external field. These included more detailed models of the fields due to magnetopause and tail currents, the latter being of great importance in interpretation of the first Mariner 10 encounter (M10-I) and the MESSENGER M1 and M2 close flybys. An explicit model of the external field allows one to extend analyses through regions of local currents (e.g., tail current sheet), accommodating nonpotential fields more readily. A number of these models (Bergan and Engle, 1981; Jackson and Beard, 1977; Ng and Beard, 1979; Whang, 1977) were applied to M10 observations and allowed for the possibility of internal field through degree 2 or 3 but only in the zonal harmonic ($m=0$) coefficient, for example, g_2^0 and g_3^0. Models allowing for axisymmetric quadrupole ($n=2$) or octupole ($n=3$) terms are characterized by a reduced dipole magnitude, relative to models allowing dipole terms only. This

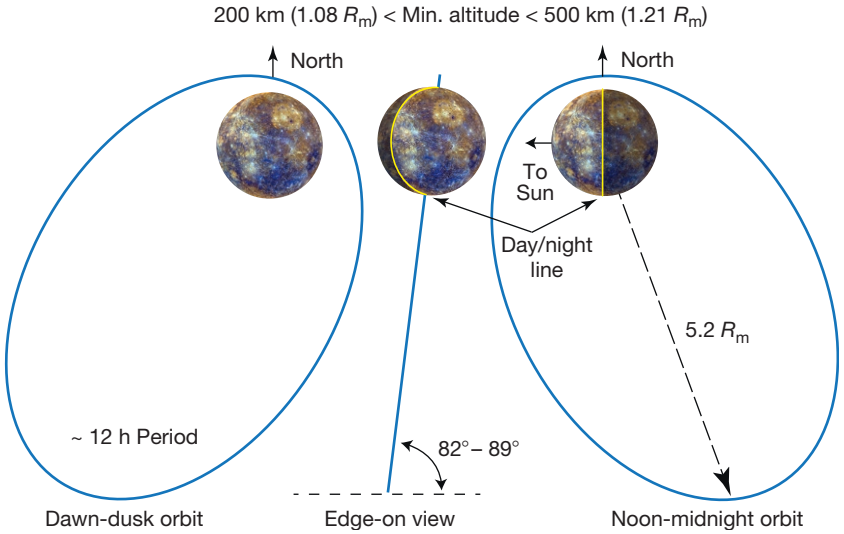


Figure 4 MESSENGER orbits about Mercury slowly evolve in local time (dawn–dusk at left; noon–midnight at right) as Mercury orbits the Sun. Closest approach to Mercury at altitudes of ~200 to ~500 km occurs at high northern latitudes (60°–74°).

Table 2 Mercury magnetic field models

Model[a]	g_1^0*(nT-*R_m^3*)*	*Data*	*Reference*	*Comments*
I1	227	I(1/2)	Ness et al. (1974a,b)	No external field modeled
I1E2	350	I(1/2)	Ness et al. (1975)	External field model improves internal fit
I1E1	342 ± 15	III	Ness et al. (1976)	III data better suited for modeling internal field
I1E1	330 ± 18	III	Ness (1979a,b)	III subsets modeled to estimate errors

[a]I1 denotes internal expansion of degree/order 1; E*n* denotes external field expansion of degree/order *n*.

applies to all such models, regardless of author or method of analysis or even the details of the explicit models chosen to represent fields of external origin. This curious effect was shown to be related to model nonuniqueness, a consequence of the limited spatial extent of the data (Connerney and Ness, 1988). With spatially limited observations, the internal field parameters and the external field parameters covary, in a predictable way, and cannot be uniquely determined, lacking additional information. Thus, following the Mariner 10 encounters, there was no evidence either for or against a nonzero axisymmetric quadrupole or octupole. Either model an offset dipole or a simple dipole (e.g., Ness et al., 1979a,b) provides an acceptable, if not entirely satisfactory, description of Mercury's internal magnetic field. This model nonuniqueness could only be resolved with additional close-in observations, preferably distributed more uniformly about the planet and in particular above the southern pole. The internal field is just sufficient to keep the dayside magnetopause above the planet's surface under nominal solar wind conditions (**Figure 5**).

Initial analyses of the MESSENGER observations used the M1 and M2 flybys, in combination with Mariner 10 observations, and a variety of explicit models of the external field, to obtain qualitatively similar results. Anderson et al. (2008, 2010) and Uno et al. (2009) found a number of dipole representations with moments ranging from 240 to 270 nT-R_m^3 and dipole tilts of 5°–12° from the rotation axis, as well as (northward) offset dipole models that fit equally well with reduced dipole moment and negligible tilt. Alexeev et al. (2010) modeled the Mariner 10 and MESSENGER flybys, fitting an elaborate external field model to each, individually, concluding that Mercury's internal field is best represented by an axial dipole, with a moment of 196 nT-R_m^3, offset from the planet's center by 405 km ($0.17R_m$) to the north. This result is rather similar to the Mariner 10 models (Bergan and Engle, 1981; Ng and Beard, 1979) characterized by an aligned dipole (with a moment of 180–207 nT-R_m^3) offset by 0.17–0.19R_m northward along the *z*-axis.

10.06.3.1.3 Discussion

MESSENGER's orbital phase has thus far provided a wealth of additional observations, well distributed in longitude and local time, but close-in observations are still limited to high northern latitudes. Thus, the nonuniqueness inherent in the limited spatial distribution of observations – the covariability of the main dipole (g_1^0) term, northward offset or g_2^0 term, and the external field – described by Connerney and Ness (1988) and Korth et al. (2004) remains extremely challenging. Only by constraining, independently, the northward offset of the dipole using relatively distant observations of the location of the magnetic equator (Anderson et al., 2011, 2012; Johnson et al., 2012) has it been possible to deduce the magnitude of the dipole (~190 nT-R_m^3, corresponding to a northward dipole offset of ~480 km) from a few Mercury years of MESSENGER orbital observations. This internal field model, along with a prescribed paraboloid magnetospheric model (Alexeev et al., 2010), fits selected MESSENGER orbits very well (Johnson et al., 2012). However, the uniqueness of this particular solution has yet to be demonstrated. The (averaged) magnetic equator locations critical to constraining the dipole offset are available at essentially two radial distances (~1.4 and ~2.5R_m) determined by MESSENGER's fixed orbit; the innermost magnetic equator observations are available at most local times but in close proximity to the magnetopause. The more distant observations are necessarily confined to a narrow range of local times (near midnight). The magnetotail current sheet geometry may be more complex than assumed, near the magnetopause, and somewhat warped across and down the tail, as is observed at Earth (Fairfield, 1980; Gosling et al., 1986; Zhang et al., 2006). Given that the external field contributes up to 40% or 50% of the field measured (Anderson et al., 2010; Johnson et al., 2012), one must be very confident in

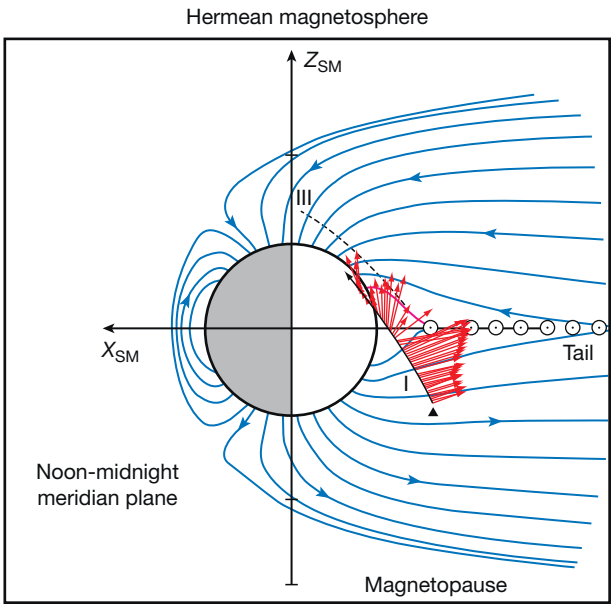


Figure 5 Magnetic field lines of a model Hermean magnetosphere (blue) in meridian plane projection (*x*–*z* plane) compared with the vector magnetic field observed along the M1 passage through the magnetotail.

knowledge of the external field in order to narrow the range of uncertainty in the internal field parameters (especially the *zonal* harmonics). The $g_1^0 - g_2^0$ covariability may only really be resolved satisfactorily after BepiColombo's well-distributed orbital measurements (**Figure 6**) are available (Scuffham et al., 2006). There remains the intriguing possibility of detecting, and characterizing, a putative crustal magnetic anomaly with the many close passes afforded by either MESSENGER or BepiColombo, but as yet no unambiguous evidence for such has emerged (Purucker et al., 2009).

10.06.3.2 Venus

Venus is similar to Earth in size ($R_v = 6051$ km), mass ($M_v = 0.815 M_e$), and density, and probably composition as well. Venus orbits the Sun once every 225 days at a mean orbital radius of 0.723 AU. Proximity to Earth has invited a relatively large number of missions, particularly in the early decades of space exploration (Colin, 1983; Russell and Vaisberg, 1983). However, unlike Earth, Venus has no measureable internal magnetic field; spacecraft have detected only weak magnetic fields associated with the solar wind interaction (Russell and Vaisberg, 1983). Venus also rotates very slowly, with a period of 243 days. The relatively slow rotation in itself does not explain the lack of a dynamo, as it is easily enough to make Coriolis forces dynamically important (Stevenson, 2003).

10.06.3.2.1 Observations

The first measurements of the magnetic field near Venus were obtained in December 1962, by the Mariner 2 spacecraft, on a flyby trajectory with a closest approach of 6.6 planet radii. Magnetic field measurements obtained by the triaxial fluxgate magnetometer (FGM) aboard Mariner 2 provided no evidence of an intrinsic planetary field (Smith et al., 1963, 1965a) at this distance. Venus could not have an internal magnetic field with a moment exceeding 0.05 that of the Earth.

The 1960s and 1970s saw the arrival of the Venera 4 (1967) and Venera 6 (1969) entry probes and the Mariner 5 (1967) and Mariner 10 (1974) spacecraft on flyby trajectories with close approach distances of 1.7 and $1.96 R_v$. These were followed by the Venera 9 and 10 orbiters in 1975 and the Pioneer Venus Orbiter (PVO) in 1978. The Veneras were placed in highly elliptical orbits with an orbital period of ~48 h, periapsis at 1545 and 1665 km altitude, respectively, and apoapsides of $\sim 18.5 R_v$. Throughout this period, the existence of an intrinsic magnetic field was debated (see Russell and Vaisberg, 1983) amid ever-decreasing estimates, or upper limits, on the Venus dipole moment. The PVO was placed in a highly elliptical orbit with an orbital period of 24 h, periapsis at 140–170 km altitude, and apoapsis at $\sim 12 R_v$. This orbit geometry afforded observation of, and a better understanding of, the interaction of the solar wind with a planetary atmosphere. These observations led to an upper limit on the Venus dipole moment of 4×10^{-5} that of Earth's (Russell et al., 1980), which was reduced further (to $\sim 1 \times 10^{-5} M_e$) by a more exhaustive analysis (Phillips and Russell, 1987) that also demonstrated a lack of evidence for magnetic fields associated with crustal remanence. The Venus Express spacecraft, in a highly elliptical polar orbit about Venus since April 2006, will not likely alter conclusions regarding the internal magnetic field of Venus (Zhang et al., 2006).

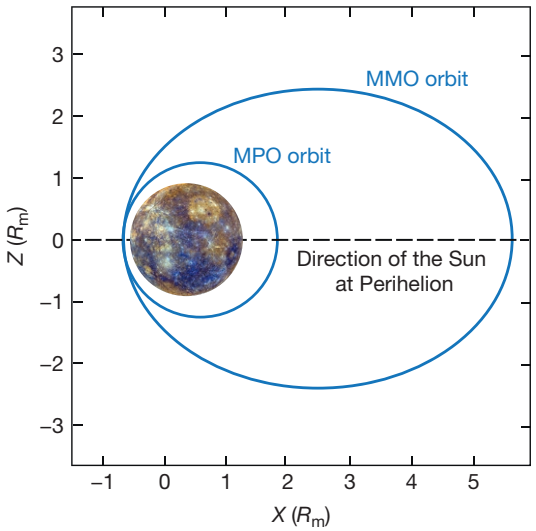


Figure 6 Coplanar polar orbits of the BepiColombo mission spacecraft. The three axis stabilized Mercury Planetary Orbiter (MPO) will complete an orbit about Mercury in about 2.3 h, while the Mercury Magnetospheric Orbiter (MMO), a spin-stabilized spacecraft, will have an orbit period of about 9.3 h; both spacecraft will have a closest approach altitude of 400 km. Launch is scheduled in July 2016.

10.06.3.2.2 Discussion

One concludes that Venus does not currently have an active dynamo; at present, little can be said of the possibility that Venus once sustained a dynamo. One would like to understand why this Earth-like planet does not have an active dynamo, and there are many possibilities (Spohn, 1991; Stevenson, 2003). The favored explanation is that sufficient convection in the fluid outer core is lacking at present, owing perhaps to the inefficiency of heat transport (relative to Earth, where plate tectonics helps cool the core). More specifically, the lack of a present-day dynamo may be the result of a cessation of plate tectonics on Venus (Nimmo, 2002), as has been proposed for Mars (Nimmo and Stevenson, 2000).

10.06.3.3 Mars

Mars is about half the size ($R_m = 3394$ km) of Earth and much less massive ($M_m = 0.108 M_e$). Mars resides at a mean orbital distance of 1.52 AU, orbiting the Sun once in 687 days. It has also been an attractive target for planetary exploration, playing host to more than 30 attempted missions since 1960 (Connerney et al., 2004). At present, one of the indefatigable Mars Exploration Rovers continues to explore the surface after more than 10 years of service (Crisp et al., 2003). The MGS spacecraft (Albee et al., 1998) fell silent in 2006 after more than 9 years in orbit, and Mars Odyssey (Saunders et al., 2004), Mars Express (Chicarro et al., 2004), and the Mars Reconnaissance Orbiter (Graf et al., 2005) all continue to acquire observations from orbit. The Mars Science Laboratory (MSL) was delivered to the surface in August 2012 to begin its search for environmental conditions suitable for past or present microbial life, and the MAVEN spacecraft is en route to Mars (arrival

21 September 2014) to begin a year-long study of the Mars atmosphere. It is the most impressive armada of spacecraft ever assembled for planetary exploration, although only one currently active vehicle – the MAVEN spacecraft – is instrumented to measure magnetic fields.

10.06.3.3.1 Observations

After five unsuccessful attempts (four by the USSR and one by the United States), Mariner 4 became the first mission to return data from Mars on 15 July 1965. Mariner 4's flyby trajectory brought the spacecraft to within $4R_m$ of the planet, from which vantage point it observed the near-Mars environment. The magnetic field was indistinguishable from the interplanetary environment; no bow shock, no magnetopause, and no magnetosphere were observed. If Mars had an intrinsic magnetic field, it had to be very small in magnitude, $<3\times10^{-4}M_e$ (M_e, the magnetic dipole moment of Earth, is 8×10^{25} G cm^3), to remain unseen (Smith et al., 1965b). This upper limit is equivalent to an equatorial surface field of 100 nT, well less than Earth's (30 000 nT) but not so much less than the as yet undiscovered magnetic field of Mercury (330 nT).

In the following three decades, numerous spacecraft visited Mars, but no spacecraft, probe, or lander instrumented to measure magnetic fields would pass close enough to the surface to establish the presence of an intrinsic magnetic field. The US missions following Mariner 4 and Mariner 9 (1971) and Viking 1 and Viking 2 (1975) were not instrumented to measure magnetic fields. Additional magnetic field observations were obtained by the Soviet Mars 2, 3, and 5 spacecraft (Ness, 1979a; Russell, 1978a, 1979) in the early 1970s and the Soviet Phobos 2 mission in 1988. These observations fueled a great deal of debate regarding the existence of a putative Mars dipole (with equatorial surface field in the range 20–65 nT), but no convincing evidence of an intrinsic field was found (Riedler et al., 1989).

Thus, it was the MGS spacecraft (Albee et al., 1998), launched in 1996 as a (partial) replacement for the unsuccessful Mars Observer, that provided the first unambiguous detection of the intrinsic magnetic field of Mars. On 11 September 1997, the MGS spacecraft was inserted into a highly elliptical polar orbit from which it would transition to a circular polar mapping orbit (nominal 400 km altitude) via a series of aerobraking maneuvers (Albee et al., 1998). Aerobraking uses atmospheric drag to slow the spacecraft, gradually reducing apoapsis by repeated passage through the atmosphere; trim maneuvers are used to raise or lower the altitude at closest approach (periapsis) to obtain the desired drag. Each aerobraking pass brought the spacecraft over the surface of Mars, to altitudes <100 km, about as close to the surface as any orbiting platform can achieve. During each aerobraking pass, the MGS magnetometer/electron reflectometer (MAG–ER) investigation (Acuña et al., 1992) acquired measurements of the vector magnetic field along the orbit track at varying altitude above the surface.

Periapsis pass 6 (day 264, 1997) provided a wealth of scientific insight in a few minutes time: lacking a significant field of global scale, Mars has no active dynamo, but must have had one in the past, when the crust acquired intense remanent magnetization (Acuña et al., 1998). These observations are illustrated in Figure 7, showing the spacecraft trajectory and

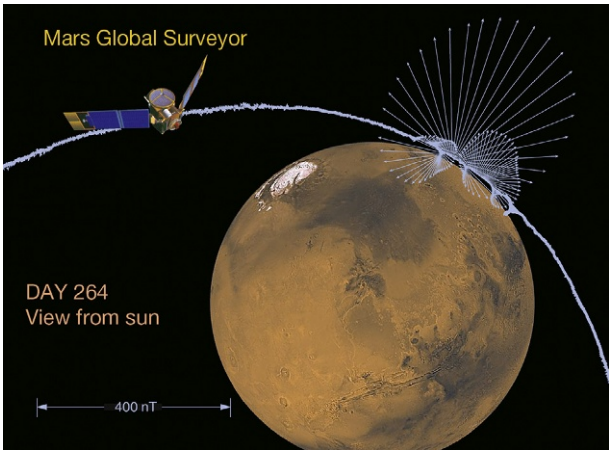


Figure 7 Mars Global Surveyor trajectory and magnetic field observations in an orbit plane projection for day 264, 1997. The vector magnetic field observed along the MGS trajectory is illustrated by vector projection onto the orbit plane; vectors originate at the spacecraft position at the time of observation. Reproduced from Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111(1–2): 1–32.

vector magnetic field measured at 3 s intervals along the trajectory. From this image, one can see that the intrinsic field is not global in scale, as it would necessarily be if it were generated deep within the planet by dynamo action. Instead, the field is localized to a small part of the trajectory, with a characteristic length scale comparable to the altitude of observation. This is the signature of a strong and localized magnetic source in the crust of Mars, just beneath the spacecraft orbit track. This particular source produces a field of nearly 400 nT at ~110 km altitude, requiring a very large volume of intensely magnetized material, with a moment of $\sim1.6\times10^{16}$ A m^2 (Acuña et al., 1998).

The MAG–ER investigation on MGS acquired magnetic field observations on over 1000 aerobraking passes during the transition to mapping orbit. On each such pass, the crustal field was sampled at low altitude along a limited ground track centered on the latitude and longitude of periapsis and extending approximately ±25° in latitude north and south of closest approach. Periapsis passes were distributed more or less randomly in longitude, as the orbit period was shortened by aerobraking, and over the course of aerobraking, the latitude of periapsis slowly evolved from ~30° north latitude to ~87° south. Thus, MGS was able to sample crustal fields over nearly the entire planet surface, which led to a map of the global distribution of magnetic sources in the Mars crust (Acuña et al., 1999). This map, reproduced here (Figure 8), is compiled by color coding, the (largely north–south) subspacecraft trajectory using the radial magnetic field component measured at that latitude and longitude. No correction for spacecraft altitude, which varies along the track from periapsis (~100 km) to 200 km, was attempted in this presentation. The most significant magnetic sources are distributed nonrandomly, largely confined to the older, heavily cratered terrain south of the dichotomy boundary; vast regions of the smooth northern plains are relatively nonmagnetic, as are the Argyre and Hellas impact craters and prominent volcanic

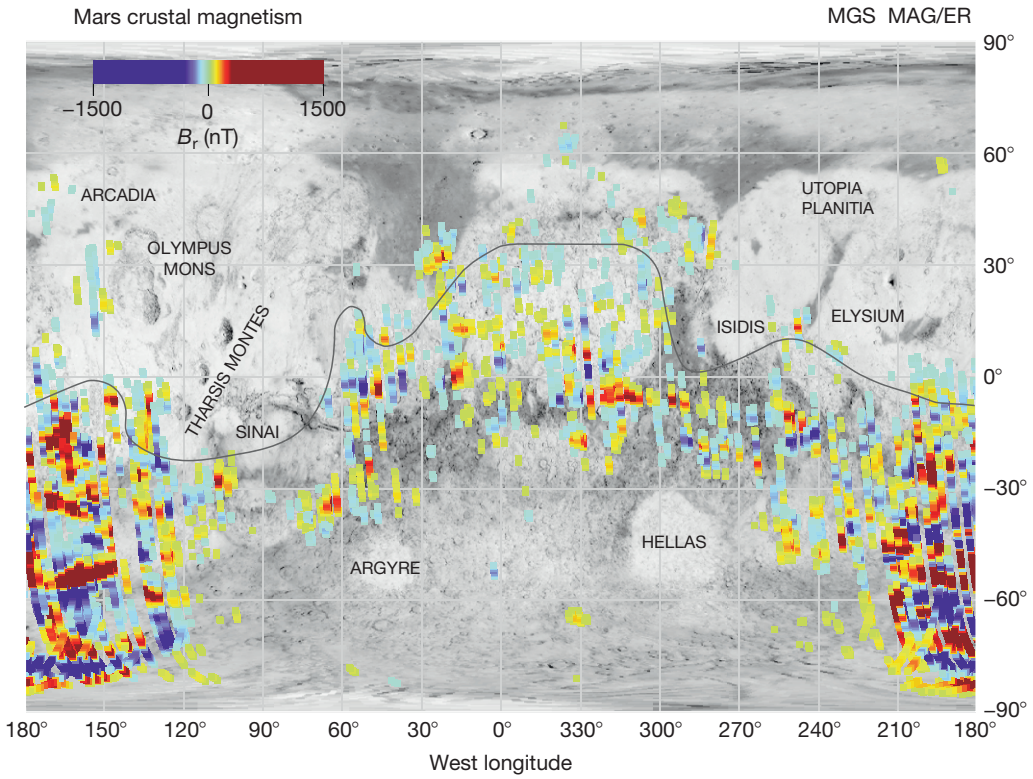


Figure 8 Global distribution of the most intense magnetic sources in the Mars crust. Magnitude of the radial magnetic field measured during the Mars Global Surveyor aerobraking passes below 200 km altitude (to ~100 km minimum altitude). Most intense sources are found in the heavily cratered southern highlands, south of and near the dichotomy boundary (thin line). Reproduced from Acuña MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.

edifices (Acuña et al., 1999). This led Acuña et al. to conclude that the Mars dynamo likely operated for a brief (few 100 My) period following accretion (~4.5 Ga) and had ceased to operate by the time of the late large impacts, about 3.9 Ga. More comprehensive studies of the magnetic fields above impact basins arrived at essentially the same conclusion (Lillis et al., 2008, 2013).

In March of 1999, MGS completed the transition to circular, polar mapping orbit with nominal altitude of 400 km, from which it would begin its mapping mission. The orbit is fixed in local time (2 a.m.–2 p.m.), and subsequent orbits cross the equator 28.6° westward of the previous orbit. Each 'mapping cycle' of 28 days duration provides uniform sampling of the globe with track-to-track separation of about 1° of longitude (59 km) at the equator. Subsequent mapping cycles increase the density of ground tracks, providing complete global coverage with track-to-track spacing of 3 km at the end of 1 Mars year (687 days). MGS passes over the same location on Mars many times; thus, the vector magnetic field at satellite altitude can be mapped with extraordinary signal to noise (Connerney et al., 2001). This particular map was compiled with observations obtained over the nightside, where time-variable fields due to the solar wind interaction are minimized, to improve the signal fidelity of the crustal field. The crustal field at mapping altitude reaches a maximum of ~220 nT over the intensely magnetized southern highlands. This and other compilations of MGS observations are available online at http://mgs-mager.gsfc.nasa.gov.

A much improved map of the field produced by crustal sources was obtained by using a more effective means to remove residual external fields (Connerney et al., 2005). This map (**Figure 9**) is a 360×180 pixel 'image' of the filtered radial magnetic field, each pixel representing the median value in a $1° \times 1°$ latitude–longitude bin. This map represents the change in radial field, or $\triangle B_r$, in nanotesla per degree latitude, as the spacecraft moves north to south over the darkened hemisphere. The map is superposed upon a Mars Orbiter Laser Altimeter (MOLA) shaded topography map (Smith et al., 1999), which appears in any pixel for which $\triangle B_r$ falls below a minimum threshold of ± 0.3 nT per degree latitude. The threshold is chosen well above the noise background, so little noise appears in the figure to obscure the context image. The extraordinary signal fidelity of this map can be appreciated by recognizing that pixels adjacent in longitude are statistically independent, by virtue of the way MGS orbits accumulate. The map spans over two orders of magnitude of signal dynamic range.

The ER instrument of the MAG–ER investigation (Acuña et al., 1992, 1998) provides another means of sensing – remotely – the magnitude of the magnetic field between the spacecraft and atmosphere (Lillis et al., 2004; Mitchell et al., 2007). The technique is similar to that used in lunar orbit to infer surface magnetic field magnitudes (Anderson et al., 1975; Lin et al., 1998; Mitchell et al., 2008) from the properties of the (reflected) electron distribution; however, at Mars, the presence of an absorbing atmosphere must be taken into account (Lillis et al., 2004). Application of this technique at Mars is

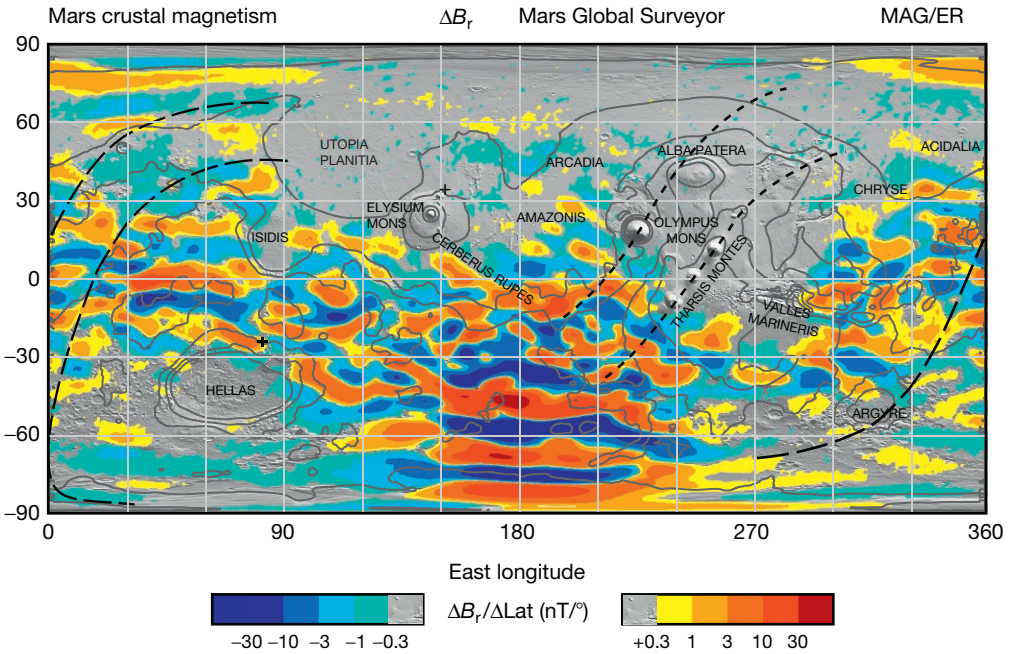


Figure 9 Map of the magnetic field of Mars observed by the Mars Global Surveyor satellite at a nominal 400 km (mapping) altitude. Each pixel is colored according to the median value of the filtered radial magnetic field component observed within the $1^\circ \times 1^\circ$ latitude–longitude range represented by the pixel. Colors are assigned in 12 steps spanning two orders of magnitude variation. Where the field falls below the minimum value, a shaded MOLA topography relief map (Smith et al., 1999) and contours of constant elevation (−4, −2, 0, 2, and 4 km elevation) are shown. Reproduced from Connerney JEP, Acuña MH, Ness NF, Kletetschka G, et al. (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102(42): 14970–14975.

particularly effective in sensing weaker fields at an altitude of about 170 km above the surface. The ER technique is, however, an indirect method of sensing the magnetic field magnitude at some distance from the spacecraft, and the application of the technique requires attention to detail. The method is based on the magnetic mirror effect and the identification of a population of electrons, sorted by pitch angle, that have reversed their motion along the magnetic field, in response to an increased field magnitude encountered en route to the surface. In the ideal case (adiabatic approximation), an electron follows a helical path along B preserving its energy and magnetic moment, $\mu = mv^2 \sin^2\alpha / 2|B|$, where α is the angle ('pitch angle') between the electron velocity v and the magnetic field. As the electron moves toward an increasing field magnitude, the electron pitch angle rotates toward 90°, at which point the electron reflects (reverses course along B) and is returned to the spacecraft. Electrons with sufficiently small pitch angle (more field-aligned) will not be reflected before reaching the surface (or being absorbed by the atmosphere), and their absence in the electron distribution allows one to identify the maximum magnetic field magnitude encountered between spacecraft and surface. The vector magnetic field measured at the spacecraft is used to infer where (beneath the orbit) the increased field magnitude was encountered and to correct the field magnitude due to crustal sources for externally applied fields. The adiabatic approximation requires that the ratio of the electron gyro radius to the characteristic dimension over which the field changes to be small and that the field does not change appreciably during a gyroperiod. In practice, one must assume that the field does not change direction appreciably between the spacecraft (where it is measured) and the electron's reflection point and that drift motion due to electric fields and magnetic gradients is negligible. Parallel electric fields, which can also reflect electrons, may be recognized by the examination of multiple electron energies (Halekas et al., 2002) and corrected for, where necessary.

A global map of field magnitudes at ~170 km altitude (Mitchell et al., 2007), with a spatial resolution of ~150 km (and using a minimum B threshold of 10 nT at 170 km altitude), compares very well with the map (**Figure 9**) compiled from magnetic field measurements. The broad regions that appear in shaded topography in **Figure 9**, in which the magnetic field falls below threshold, are nearly identical to the weak or zero field regions (to an accuracy of a few nanoteslas) mapped by the ER (Lillis et al., 2008; Mitchell et al., 2007). The ER method works best where field lines are mostly radial at satellite altitude, as over a crustal field well approximated by a radial dipole. The method will return either no result or a less well-constrained result above field lines that are essentially horizontal beneath the spacecraft. The ER method was originally developed to sense magnetic fields above the lunar crust, populated by weak sources of relatively small spatial scale or coherence. In that environment, the field measured at satellite altitude may be straight line continued nearly all the way to the surface, due to the limited reach of the weak crustal sources. At Mars, crustal fields are much stronger and of large spatial scale, limiting the accuracy with which the field may be continued (via the straight-line approximation, i.e., nonself consistently) down toward the reflection point.

10.06.3.3.2 Models

The representations routinely employed throughout the solar system – dipoles and spherical harmonic expansions – lose much of their appeal when applied to the crustal field on Mars. Such global models require an awkward number of model parameters or coefficients to represent the crustal field. For example, Cain's (Cain et al., 2003) spherical harmonic model extends to $N_{max}=90$, requiring $(N_{max}+1)^2-1=8280$ model coefficients to calculate the field at one point above the surface. Likewise, Purucker's many-dipole model (Purucker et al., 2000) uses 11 500 dipoles distributed on a spherical surface ($1R_m$) to represent the field by superposition. There are a variety of spherical harmonic (Arkani-Hamed, 2001a, 2002; Cain et al., 2002) and equivalent source models (Langlais et al., 2004; Purucker et al., 2000) to chose from. These models can be obtained from the authors in electronic form. They offer a means of extrapolating the field from the measurement surface (400 km) and volume spanned by aerobraking passes to nearby points; however, none can be expected to accurately predict the field near the source region.

Mars crustal magnetism is amenable to study using many of the methods and tools developed for interpretation of magnetic surveys on Earth. In *source modeling*, one attempts to fit the observed vector field using one or more magnetized sources, for example, a collection of thin plates and prisms, often guided by constraints suggested by local geology and material properties. General techniques for the *continuation* of a potential field (e.g., Blakely, 1995) may also be applied to extrapolate the field away from a surface upon which the vector field is known. Downward continuation (toward the source) must be done with caution, as it is essentially a differencing operation that amplifies short-wavelength random noise (exponentially in z). Source modeling must also be done with caution, as it is not possible to uniquely determine crustal magnetizations using measurements external to the source.

Connerney et al. (1999) used a source model to fit observations obtained during several aerobraking passes over the intensely magnetized southern highlands (Terra Cimmeria and Terra Sirenum). These quasiparallel features appear lineated in the east–west (cross track) direction; many can be traced over 1000 km in length. The model consisted of multiple uniformly magnetized parallel thin plates, aligned (and extending to infinity) in the east–west direction. The volume magnetization of each strip was determined using a linear inverse methodology to best fit the vector observations along one or more aerobraking passes (**Figure 10**). The observations could be fit well with a model characterized by strips of alternating polarity and volume magnetizations of as much as $\pm 20\ A\ m^{-1}$ (assuming a 30 km thick plate). This figure is likely a lower bound, since only the product of thickness and volume magnetization is constrained by the observations (Connerney et al., 1999). If the magnetization is borne in a thin layer, say 3 km, then volume magnetizations of order $100\ A\ m^{-1}$ would be required. This is a rather astonishing intensity of magnetization required throughout large volumes of crust. At the other extreme, Parker (2003) developed a theory to establish the minimum possible intensity of magnetization consistent with the data modeled by Connerney et al. Interpolating his results to a crustal thickness of 30 km, he found a minimum magnetization intensity of $\sim 10\ A\ m^{-1}$ is required to satisfy MGS observations. Connerney et al. (1999) found the magnetic lineations on Mars sufficiently reminiscent of the pattern of magnetization associated with seafloor spreading on Earth that they proposed a similar mechanism – crustal spreading in the presence of a reversing dynamo – for the evolution of the Mars crust.

Sprenke and Baker (2000) modeled crustal magnetization in the southern highlands using the same aerobraking observations and inverse methods used by Connerney et al. (1999). Their models imposed constraints on the direction of magnetization ('normal' and 'reversed') and allowed for a variation in the intensity of magnetization in the cross track direction. They found similar magnetization intensities and obtained a good fit to the observations with a seafloor spreading model, allowing for variation of magnetization intensity (but not direction) along the stripes. The more uniformly distributed observations obtained at mapping altitude over the same region were downward continued using the Fourier transform method (Jurdy and Stefanick, 2004). Continuation to a 100 km altitude surface allowed a favorable comparison with the aerobraking observations and revealed abrupt lateral changes in field direction and magnitude suggestive of, but not requiring, reversed lineations.

There is great interest in identifying a Mars paleopole, and a number of authors have sought to determine ancient pole position(s) on Mars by fitting various magnetized sources to MGS magnetic field observations (Arkani-Hamed, 2001b; Arkani-Hamed and Boutin, 2004; Boutin and Arkani-Hamed, 2006; Frawley and Taylor, 2004; Hood and Zakharian, 2001; Hood et al., 2010; Langlais and Purucker, 2007; Sprenke and Baker, 2000). A large number of putative paleopoles have been proposed but as yet no consensus has emerged (for a recent compilation, see Milbury et al., 2012). A general consideration of nonuniqueness in the interpretation of potential fields would suggest that without additional constraints, the problem remains intractable (Connerney et al., 2004). In one case where a priori information (a gravity high) may help constrain the location of a magnetic source (one associated with the Apollinaris volcanic edifice), a paleopole location (65° S, 59° E) near the current Mars rotational pole was found by two sets of authors (Hood et al., 2010; Langlais and Purucker, 2007). However, Apollinaris was active a few hundred million years *after* the inferred demise of the dynamo, based on the analysis of the giant impacts (Lillis et al., 2013), which ought to produce a more dependable record of dynamo history. Where the association between volcanism and magnetic field is greatest, volcanism has either completely or partially demagnetized a preexisting magnetic imprint (Connerney et al., 2005); long-lived volcanoes, in bulk and sensed from orbit, are unlikely to be effective recorders of paleofield directions (e.g., Lillis et al., 2013). An alternative approach to locating paleopoles, based on a statistical analysis of many sources (Sprenke, 2005), suggests a preference for paleopoles near 17° N, 230° E, and 17° S, 50° E, near those proposed earlier based on geomorphology and analysis of grazing impacts (Schultz and Lutz, 1988). In another statistical approach, Milbury and Schubert (2010) searched among a distribution of paleopoles, assigning a magnetization intensity and direction to portions of the crust, in such a way as to best match the observed magnetic field, taken, for convenience, from several spherical harmonic model representations. They

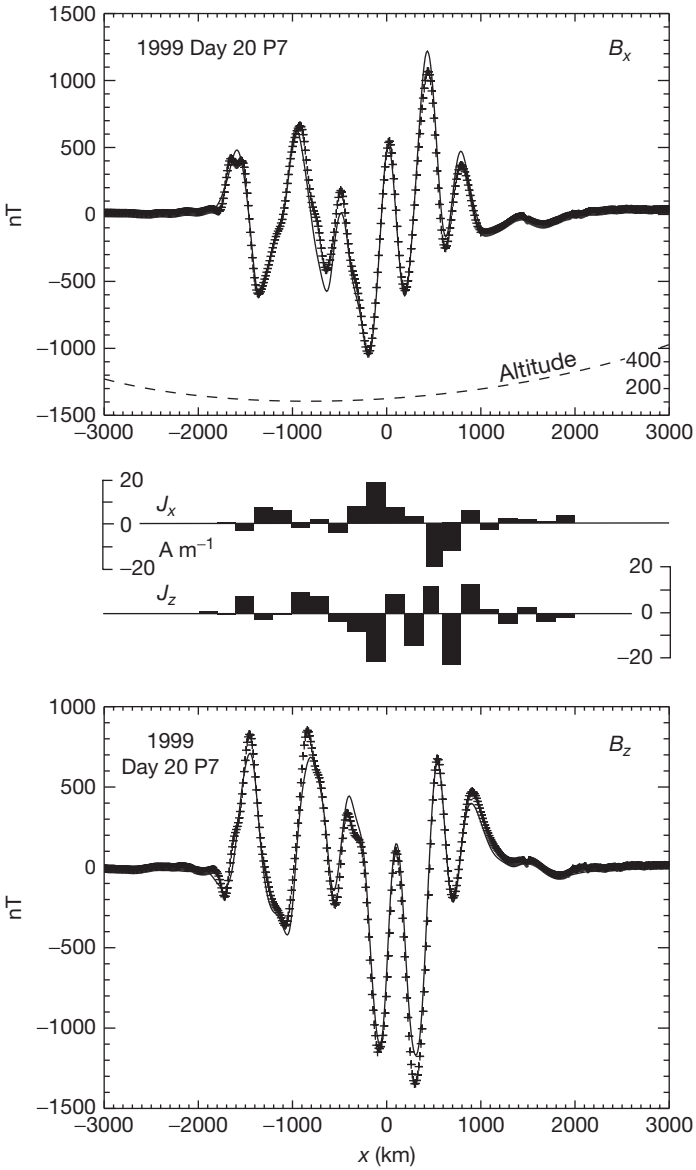


Figure 10 Vector magnetic field measured by MGS during aerobraking on calendar day 20, 1999. Periapsis occurred at 68.0° S, 181.2° W and 106 km altitude. The x (north) and z (down) components of the vector field sampled every 3 s (crosses) are compared with a model fit (solid line). Altitude variation in kilometers indicated with a dashed line. Observations plotted as a function of distance x north and south of an origin at 53° S and at the longitude of periapsis. Model consists of 20 uniformly magnetized slabs aligned with the y-axis (east–west) and infinite in extent along y. The x and z components of the model crustal magnetization per unit volume (A m^{-1}) are indicated in the bar graphs between the two panels. Reproduced from Connerney JEP, Acuña MH, Wasilewski PJ, Ness NF, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.

found a weak but significant correlation for spherical harmonic degrees 2 and 3 and a set of paleopoles and none for higher degrees. This may indicate that only a global-scale correspondence exists, but it may also reflect a sampling bias, in that large portions of the crust (northern lowlands, Tharsis complex, and major impact basins) were removed from the model. The paleopoles are found at low latitudes to midlatitudes and are suggestive of polar wander and polarity reversal (Milbury and Schubert, 2010). Coexistence of near-equatorial and near-polar paleopoles can be reconciled with the expectation of a moderate angular separation of dipole (magnetic) and rotational axes if one invokes significant polar wander during the active dynamo era (e.g., Boutin and Arkani-Hamed, 2006; Milbury et al., 2012; Sprenke, 2005).

10.06.3.3.3 Discussion

Mars has no global field and therefore no dynamo at present, but must have had one in the past when the crust acquired intense remanent magnetization. It is likely that a molten iron core formed early, after or during hot accretion 4.5–4.6 Ga, and for at least a few hundred million years, a substantial global field was generated by dynamo action in the core. The chronology proposed by Acuña et al. (1999) attributes the global distribution of magnetization to an early demise of

the dynamo, prior to the last great impacts (~4 Ga) that left large unmagnetized basins in the crust. This view has been supported by more complete analyses of the large impact basins (Lillis et al., 2008, 2013), leading to more precise estimates of the dynamo's demise (activity extinguished before formation of Hellas and Utopia basins 4.0–4.1 Ga). Schubert et al. (2000) argued in favor of a late onset of the dynamo and proposed that the southern highlands crust acquired magnetization from local heating and cooling events that postdate the era of large impacts and basin formation. Early onset and cessation of the dynamo are difficult to reconcile with the notion of a dynamo driven by solidification of an inner core (Schubert et al., 1992), the preferred energy source for the Earth's dynamo. Alternatively, an early dynamo can be driven by thermal convection, with or without plate tectonics, for the first 0.5–1 Gy (Breuer and Spohn, 2003; Connerney et al., 2004; Schubert and Spohn, 1990; Stevenson et al., 1983), persisting as long as the core heat flow remains above a critical threshold for thermal convection (Nimmo and Stevenson, 2000).

After more than 2 full years of mapping operations, MGS produced an unprecedented global map of magnetic fields produced by remanent magnetism in the crust (Connerney et al., 2005). This map (**Figure 9**) reveals contrasts in magnetization that appear in association with known faults, variations in magnetization that are clearly associated with volcanic provinces, and magnetic field patterns that appear shifted along small circles in the manner of transform faults at spreading centers (Connerney et al., 2005). Connerney et al. proposed that the entire crust acquired a magnetic imprint via spreading in the presence of a reversing dynamo and that erasure of this imprint occurred where the crust was buried (thermal demagnetization) by flood basalts to depths of a few kilometers. In the remaining magnetic imprint, one can identify the telltale signature of transform faults, in Meridiani Planum, and the signature of craters that formed both before and after the demise of the dynamo. Of course, transform faults are unique to plate tectonics, so if these features are indeed transform faults, then the Mars crust formed via seafloor spreading as on Earth (Connerney et al., 1999; Sleep, 1990). The interpretation of this map is not without controversy, and many regard the seafloor spreading analogy championed by Connerney and colleagues as speculative, in large part due to the paucity of correlative observations. The alignment of the great volcanic edifices on Mars is consistent with plate motion over a mantle plume (Connerney et al., 2005) or, conversely, volcanic chains formed above subducting slabs (Sleep, 1990, 1994; Yin, 2012, personal communication). The topographical relief along much of the dichotomy boundary has been interpreted as a series of ridge/transform fault segments (Sleep, 1994). A recent structural analysis of the Valles Marineris fault zone (Yin, 2012) likens this trough system to the left-slip, transtensional Dead Sea fault zone on Earth joining the Red Sea spreading center to the Zagros suture zone in the north: an undisputed plate boundary. It is difficult to understand how such a structure evolved on Mars in the absence of plate tectonics.

Fairen et al. (2002) proposed that the magnetic lineations might instead be the signature of accretionary terrains formed during an early era of plate tectonics, while others have proposed the formation of a series of quasiparallel dike intrusions as alternatives to the crustal spreading hypothesis (Hood et al., 2007; Nimmo, 2000; Ravat, 2011). Using an automated cluster analysis of lineated features in Connerney et al.'s (2005) map, Kobayashi and Sprenke (2010) identified two groups, or clusters, of features that they attributed to thermoremanent magnetization of chains of intrusions formed by lithospheric drift over mantle hot spots. However, none of these proposed mechanisms manufacture magnetic material with the uniformity, efficiency, and spatial scale of crustal spreading in the presence of a global magnetic field. To account for the enormous magnitude of the field above the Mars crust, one requires a great deal of material coherently magnetized with a rather startling intensity of magnetization. It is difficult to imagine any process that can compete with seafloor spreading in this regard.

Magnetic mapping is an even more powerful tool on Mars than on Earth, where it has been indispensible in understanding the evolution of the Earth, especially in the development of the unifying theory of plate tectonics. On Earth, magnetic surveys are complicated by the presence of the main field, which makes measurement of crustal anomalies challenging: the Earth's magnetic anomalies are very weak, compared to the main field, resulting in limited signal to noise in the measurement. Magnetic anomalies are also more difficult to interpret on Earth because they may result from either induced or remanent magnetization, or both. Mars, lacking a main field, is free of such complications; the global map of its crustal field is quite remarkable, affording an opportunity to apply the methods of exploration geophysics on a global scale. It is clear that the crust of Mars retains a magnetic imprint of its formation and subsequent evolution. That history may be read with increasing resolution as magnetic surveys are conducted at lower altitudes. One may anticipate great insights from such surveys as well as from Mars paleomagnetism when that field develops.

10.06.4 Gas Giants

10.06.4.1 Jupiter

Jupiter, the largest ($R_j = 71\,372$ km) and most massive ($M_j = 318 M_e$) planet in the solar system, resides at a mean orbital distance of 5.2 AU from the Sun. Jupiter, capturing much of the gas component of the solar nebula, is largely a solar mix of H and He, possibly with a rock and ice core of some tens of Earth masses. The Jovian system has been visited often by spacecraft, both as a destination and as a waypoint for vessels destined to travel onward. A passing spacecraft can use Jupiter's considerable mass and orbital velocity to advantage, acquiring a boost to greater radial distances (e.g., Cassini and New Horizons to Pluto) or to achieve higher solar latitudes (e.g., Ulysses), than would otherwise be possible. Thus, Jupiter's orbit serves as a crossroads for travelers to other worlds as well.

Jupiter was known to have a magnetic field long before the first *in situ* observations were obtained. Burke and Franklin (1955) had detected nonthermal decameter wavelength emissions (22 MHz) from Jupiter, leading to speculation that Jupiter possessed a magnetic field of internal origin. Observations at decimetric (~1 GHz) wavelengths subsequently revealed synchrotron radiation, emitted by high-energy electrons trapped in a Jovian Van Allen belt. This emission, as observed, is modulated by the rotation period of the planet (9.925 h). Since synchrotron radiation is narrowly beamed in a direction

perpendicular to the magnetic field guiding the electron motion, the modulation can be used to infer the geometric properties of the field. Thus, the tilt (9.5°) and phase (200° λ_{III}) of the Jovian dipole were deduced (Morris and Berge, 1962), and estimates of the dipole offset (from the origin) were made (Warwick, 1963), well before the first spacecraft arrived at Jupiter. The magnitude of the field, however, remained in doubt.

Continued observations of Jovian radio emission at decameter wavelength revealed yet another periodicity in the occurrence of radio storms, one that matched the orbital period (42.46 h) of the innermost Galilean satellite, Io (Bigg, 1964). The intensity and occurrence rate of radio storms vary dramatically with Io's orbital phase, as well as Jovian central meridian longitude. Activity peaks when Io's orbital phase is near 90° and 240° (measured in the direction of Io's orbital motion from geocentric superior conjunction). Io's influence on Jovian radio emissions was attributed to an electrodynamic interaction (Goldreich and Lynden-Beil, 1969) that drives currents along magnetic field lines to and from Jupiter's magnetosphere.

10.06.4.1.1 Observations

Pioneer 10, the first spacecraft to enter the Jovian magnetosphere, passed within 2.8R_j of the planet at closest approach in December 1973. Pioneer 10's flyby trajectory carried it through the Jovian system at relatively low magnetic latitudes (Figure 11). This region, near the equator, is home to a large-scale system of azimuthal currents (ring currents) extending from about 5R_j to the outer reaches of the magnetosphere (~100R_j). This current system imparts a disklike geometry to the magnetic field, with field lines at great distances stretched nearly parallel to the equator (leading to the descriptive 'Jovian magnetodisk'). Precisely 1 year later, Pioneer 11 approached to within 1.6R_j on a retrograde λ_{III} trajectory. The Pioneer 11 trajectory provided measurements over a wide range of magnetic latitudes (Figure 11) and longitudes.

The Pioneer 10 and 11 spacecrafts were instrumented to measure magnetic fields with a vector helium magnetometer (VHM) provided by the Jet Propulsion Laboratory (Smith et al., 1975b). This instrument could measure a maximum of ±1.4 G in the highest of its eight dynamic ranges. Pioneer 11 also carried a small high-field triaxial FGM provided by the Goddard Space Flight Center (Acuña and Ness et al., 1976). The FGM was added as a precaution, to extend measurement capability to a field of 10 G along each axis (the Earth's field magnitude ~0.6 G at the pole). The magnitude of Jupiter's field was unknown at the time of launch, and the concern was that without a high-field instrument, Pioneer 11 might not have the dynamic range needed to make observations throughout the encounter. Fortunately, the maximum field experienced by Pioneer 11 was ~1.12 G, observed just as the spacecraft went into occultation on its way to closest approach at 1.6R_j. There are small but significant differences between the two data sets that have not been fully resolved, so most analyses have tended to use data from one or the other but not both.

The two Voyager encounters followed in March and July of 1979. Like Pioneer 10, the Voyagers remained near the Jovigraphic equator, confined to relatively low magnetic latitudes. Voyager 1 was targeted for a close flyby of Io and the Io flux

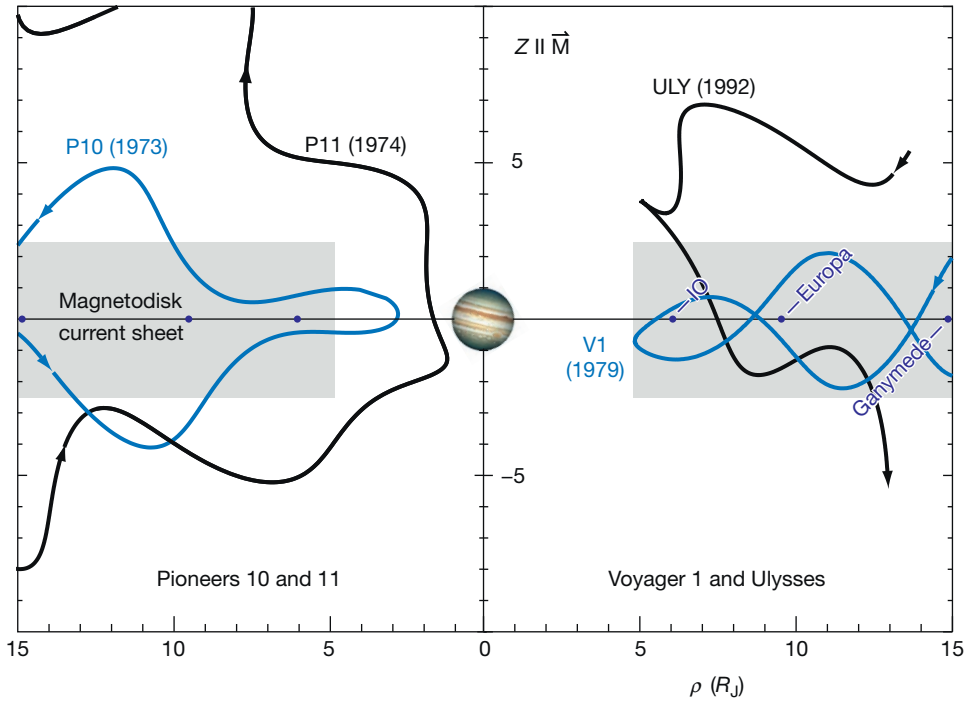


Figure 11 Spacecraft flybys (Pioneer 10 and 11, Voyager 1, and Ulysses) of Jupiter in a magnetic equatorial (cylindrical) coordinate system. The z-axis is aligned with Jupiter's magnetic dipole axis. The spacecraft distance from the magnetic equator (z) as a function of distance from the z-axis (ρ) is given in units of Jupiter radius. The Galilean satellites Io, Europa, and Ganymede trace out arcs (not shown) in this coordinate system crossing the equator at the radial distances indicated. The shaded region approximates the washer-shaped system of azimuthal ring currents that gives rise to the magnetodisk configuration of the Jovian magnetosphere.

tube (IFT); its closest approach to Jupiter was $4.9R_j$, just inside the orbit of Io ($5.95R_j$). The maximum field measured by Voyager 1 was just ~3330 nT (Ness et al., 1981). The Voyager 2 flyby was relatively distant, with a closest approach distance of $10R_j$ (Ness et al., 1982). In this region of the magnetosphere, the measured magnetic field has a significant contribution due to the magnetodisk currents. These will tend to reduce the θ component of the planetary field near the equator and introduce a component of the field in the radial direction that changes sign in crossing the magnetic equator. Local currents cannot be modeled with potential fields and, if present, may lead to large errors in estimates of the planetary field if not accounted for (Connerney, 1981).

Jupiter was not visited again until the Ulysses flyby on 8 February 1992, more than two decades after the Voyager flybys (Balogh et al., 1992). Ulysses was designed to study the heliosphere at high solar latitudes; the Jupiter flyby was used to kick the spacecraft into a high-inclination orbit about the Sun. The spacecraft approached Jupiter at relatively high northern latitude, measuring a maximum field of 2372 nT just after closest approach at $6.3R_j$. This encounter was much like that of Voyager 1, in that for much of the time near closest approach, the spacecraft was immersed in the local currents of the Jovian magnetodisk.

The recently completed Galileo mission to the Jovian system was designed to deliver an atmospheric probe, explore the distant magnetosphere, and provide opportunities for many satellite encounters (Johnson et al., 1992). Galileo's mission plan was not well suited to provide appreciable additional information on the planetary field. Likewise, the Cassini spacecraft encounter was a distant flyby, engineered to speed the spacecraft onward to its destination (Saturn) in the outer solar system. The spacecraft that passed relatively close to Jupiter did so once, providing *in situ* observations along but one path; the spacecraft (Galileo) that enjoyed multiple orbits remained at a distance and near the equatorial plane. In order to characterize Jupiter's magnetic field, and the dynamo that generates it, observations must be obtained at close-in radial distances (where higher-degree and higher-order harmonics of the field are relatively large) and well distributed globally about the planet. The Juno spacecraft launched in August 2011 and now en route to Jupiter for a July 2016 arrival (Bolton et al., 2010) will do just that. Juno's mission plan wraps the planet in a net of observations (**Figure 12**) provided by 32 near-polar orbits with periapsis just above the cloud tops, uniformly distributed in longitude (separation of 12°). The Juno magnetic field investigation is designed to map Jupiter's magnetic field with very high accuracy, sufficient to characterize the field to degree and order 12–15, depending on the dynamo core radius (estimated $0.75–0.85R_j$) and other assumptions (Connerney et al., 2014).

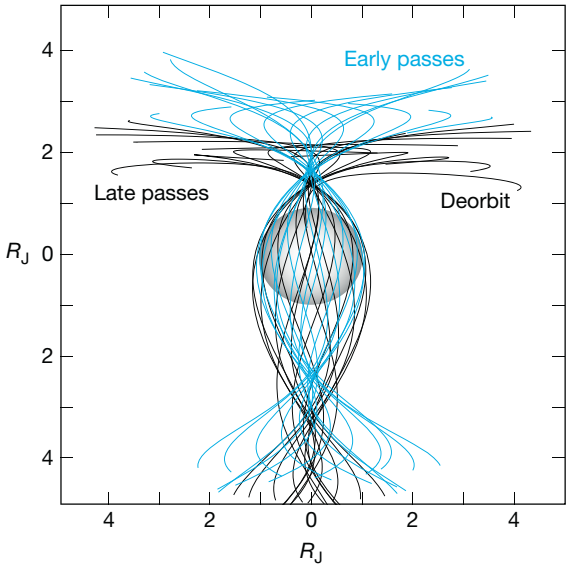


Figure 12 Orthographic projection of the near Jupiter ($r < 5R_j$) portion of Juno's 11-day science orbits. Orbits 2–16 (blue) are acquired in the first 6 months of mapping and are equally spaced in longitude (at closest approach) by 24°. Orbits 17–31 (black) are acquired after a trajectory correction that places these passes between those already acquired, resulting in complete global coverage with longitude spacing of ~12°. Juno's insertion into orbit about Jupiter (JOI) occurs in July 2016.

10.06.4.1.2 Models

The large number of Jovian magnetic field models available is a result of several factors, interest in the Jovian system undoubtedly foremost among them. There is a wealth of observations of Jovian phenomena, many of which (e.g., radio emission and auroras) depend critically on details of the magnetic field near the planet where few direct observations exist. The need is great. In addition, techniques for modeling the field have evolved, and improved, in part in response to the challenges of separating the planetary field from that produced by external currents. Techniques have also evolved to take advantage of new observations (e.g., emission from the foot of the IFT) that serve as valuable constraints on the field near the surface. Here, we describe a subset of the available models, intended to illustrate the development of new approaches.

The first models of Jupiter's planetary field were the OTD approximations (D2 and D4) obtained from the Pioneer 10 and 11 VHM observations (Smith et al., 1976). Analyses of the Pioneer 11 FGM observations suggested that a more capable parameterization was needed, leading to development of the O_4 model (Acuña and Ness, 1976), the first of many spherical harmonic models of Jupiter's internal field. This model and the spherical harmonic analysis (SHA) model of Smith et al. (1976) allowed spherical harmonics of degree and order 3 (octupole) for the internal field and approximated external fields with external harmonics of degree 1 or 2. **Table 3** gathers together the spherical harmonic coefficients of several models, fit to various subsets of data and using a variety of methods. Estimated errors are given for some of these models, but not reproduced here, for brevity. The original publication should be consulted.

The Pioneer 11 VHM and FGM observations have proved most useful for studies of the planetary field, in large part due to the relatively good spatial distribution of observations afforded by that spacecraft's trajectory. The problem of inverting flyby observations to obtain estimates of the spherical harmonic coefficients has been posed as an eigenvalue problem and solved using the SVD method (Connerney, 1981). This technique allows partial solutions to underconstrained linear inverse problems and can be used to show how even

Table 3 Jovian magnetic field models: spherical harmonics

Epoch		*Voyager 1 and IFT*	*Voyager 1, IFT, Pioneer*	*1992.1: Ulysses*	*1979.2: Voyager 1*	*1974.9: Pioneer 11*		
Coefficient number	*Coefficient*	*VIT 4*	*VIP 4*	*Ulysses 17ev*	*V1-17ev*	O_6	O_4	*SHA*
1	g_1^0	428077.0	420543.0	410879.0	420825.0	424202.0	421800.0	409200.0
2	g_1^1	−75306.0	−65920.0	−67885.0	−65980.0	−65929.0	−66400.0	−70500.0
3	h_1^1	24616.0	24992.0	22881.0	26122.0	24116.0	26400.0	23100.0
4	g_2^0	−4283.0	−5118.0	7086.0	−3411.0	−2181.0	−20300.0	−3300.0
5	g_2^1	−59426.0	−61904.0	−64371.0	−75856.0	−71106.0	−73500.0	−69900.0
6	g_2^2	44386.0	49690.0	46437.0	48321.0	48714.0	51300.0	53700.0
7	h_2^1	−50154.0	−36052.0	−30924.0	−29424.0	−40304.0	−46900.0	−53100.0
8	h_2^2	38452.0	5250.0	13288.0	10704.0	7179.0	8800.0	7400.0
9	g_3^0	8906.0	−1576.0	−5104.0	2153.0	7565.0	−23300.0	−11300.0
10	g_3^1	−21447.0	−52036.0	−15682.0	−3295.0	−15493.0	−7600.0	−58500.0
11	g_3^2	21130.0	24386.0	25148.0	26315.0	19775.0	16800.0	28300.0
12	g_3^3	−1190.0	−17597.0	−4253.0	−6905.0	−17958.0	−23100.0	6700.0
13	h_3^1	−17187.0	−8804.0	−15040.0	8883.0	−38824.0	−58000.0	−42300.0
14	h_3^2	40667.0	40829.0	45743.0	69538.0	34243.0	48700.0	12000.0
15	h_3^3	−35263.0	−31586.0	−21705.0	−24718.0	−22439.0	−29400.0	−17 100
16	g_4^0	−22925.0	−16758.0					
17	g_4^1	18940.0	22210.0					
18	g_4^2	−3851.0	−6074.0					
19	g_4^3	9926.0	−20243.0					
20	g_4^4	1271.0	6643.0					
21	h_4^1	16088.0	7557.0					
22	h_4^2	11807.0	40411.0					
23	h_4^3	6195.0	−16597.0					
24	h_4^4	12641.0	3866.0					
Magnetodisc								
	R_0	5.0	5.0	7.1	5.0 (UR)			
	R_1	56.0	56.0	128.0 (UR)	56.0			
	D	3.1	3.1	3.3	3.1			
	$\mu_0 I_0/2$	185.0	185.0	137.0	185.0			
	Θ_0	6.5	6.5	8.2	6.5			
	Φ_0	206.0	206.0	200.0	206.0			

Schmidt normalized spherical harmonic coefficient in nT, referenced to Jupiter system III (1965) coordinates, and $1R_j = 71\,398$ km for Ulysses; $1R_j = 71\,323$ km for Voyager 1.
Voyager 1 17ev model from Connerney et al., 1982a,b. Models VIP4 and VIT4 used the magnetodisc model fitted to Voyager 1 observations (V1-17ev) as fixed parameters.
The notation 'UR' refers to unresolved parameters. Pioneer 11 O_4 model coefficients as tabulated for system III (1965) by Acuña et al., 1983a,b (originally (1957 system III) from Acuña and Ness, 1976).
Pioneer 11 SHA model originally (1957 System III) from Smith et al. (1976).

small measurement errors – such as noise or unmodeled external fields – can lead to large errors among the internal field coefficients. Pioneer 11 also remained largely outside of the region of current flow (equatorial magnetodisk) near closest approach, so external fields are less of a problem.

The Voyager 1 observations yielded a considerably reduced dipole moment, upon preliminary analysis, which was attributed to the presence of magnetodisk currents that were not adequately modeled (Ness et al., 1979a). Subsequent analyses of the Voyager 1 observations used an explicit model of the magnetodisk (Connerney et al., 1981) in combination with a spherical harmonic model for the planetary field. An inversion scheme that allowed for the variation of the parameters of the magnetodisk along with those of the internal field resulted in a partial octupole model (V1 17ev) of the planetary magnetic field (Connerney et al., 1982a). The Voyager 1 model of the field (epoch 1979) could be compared with the Pioneer 11 model (GSFC O_4) of the field (epoch 1973) to limit Jovian magnetic secular variation (g_1^0) to no more than 0.2% per year (Connerney and Acuña, 1982). By comparison, the present-day decrease in the terrestrial g_1^0 term is about 0.075% per year.

Preliminary analysis of the Ulysses encounter observations (Balogh et al., 1992) suggested good agreement with the magnetic field calculated using the O_6 planetary field augmented with a magnetodisk current sheet model (Connerney et al., 1981) if one allowed for variations of the external field during the encounter. The Ulysses close approach observations were obtained in the immediate vicinity of the inner edge of the current sheet, so some attention to external fields is required. Dougherty et al. (1996) found the data consistent with Connerney's O_6 model, allowing for a reduction of the g_1^0 coefficient (main dipole term), from 4.242 to 4.059 G. Connerney et al. (1996) derived an octupole model of the field (Ulysses 17ev), also with a reduced dipole g_1^0 coefficient (4.109 G), and demonstrated that the internal zonal harmonic coefficients covary with the parameters of the current disk, complicating attempts

to separate internal and external fields. The magnetodisk current system was considerably weaker during the Ulysses epoch, relative to that observed during the Pioneer and Voyager epochs. Connerney et al. (1996) quoted a figure of 20% less total integrated magnetodisk current at the time of Ulysses encounter, Dougherty et al. (1996) quoting a 28% reduction. Differences among the Ulysses era (Ulysses 17ev) model and those of Voyager (Voyager 1 17ev) and Pioneer (O_4) appear within, or comparable to, the estimated parameter errors (Connerney et al., 1996); so there is little evidence of secular variation of the main field from these data (Connerney and Acuña, 1982; Dougherty et al., 1996) or from the more distant observations acquired throughout Galileo's tour (Russell and Dougherty, 2010; Yu et al., 2009).

The O_6 model of Jupiter's magnetic field used observations from both Pioneer 11 (1973) and Voyager 1 (1979), assuming that the magnetic field has not changed appreciably between encounters and adds a constraint on Jupiter's harmonic spectrum (Connerney, 1992). This model was designed to benefit from the better spatial distribution of observations afforded by combining trajectories. It was also designed to allow for the possibility that unmodeled fields of higher degree and order than octupole might be present in the observations (particularly those of Pioneer 11, passing to within $1.6R_j$). This model represents a partial solution to an underdetermined inverse problem (spherical harmonic expansion to degree and order 6) combined with an explicit model of the magnetodisk. Another distinguishing feature of this model was the application of a weight $(r_c/a)^{n+2}$ to parameters of degree n, where r_c, the core radius, was taken to be $0.7R_j$. This weighting encourages a model solution with terms of degree n contributing equally to the mean-squared field amplitude on the (presumed) core boundary, as is observed for Earth, through degree 12 or so (Langel, 1987). The Schmidt coefficients of this model ('O_6' for octupole part of a degree 6 expansion) are listed in **Table 3**. The O_6 model is characterized by smaller quadrupole and octupole moments, relative to the comparable O_4 model of Acuña and Ness (1976) and in that regard is more Earth-like in harmonic content (**Figure 13**).

The discovery of infrared emission at the foot of the IFT in Jupiter's ionosphere (Connerney et al., 1993) provided an entirely new and extremely valuable constraint on magnetic field models. This emission occurs in Jupiter's polar ionosphere where field lines threading the satellite Io (or its conducting ionosphere) at an orbital distance of $5.95R_j$ intersect the Jovian ionosphere. The feature is observed to move across the disk of Jupiter in concert with Io's orbital motion, tracing a path around each magnetic pole (**Figure 14**). This path serves as a fiducial marker, providing an unambiguous reference on

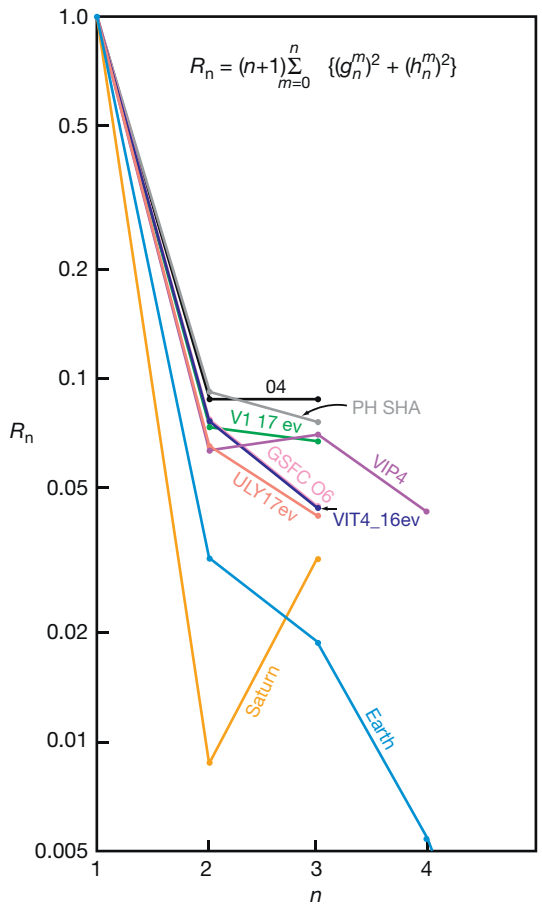


Figure 13 Relative harmonic content (Lowes' spectrum) of spherical harmonic models of the Jovian field, compared with that of Earth and Saturn.

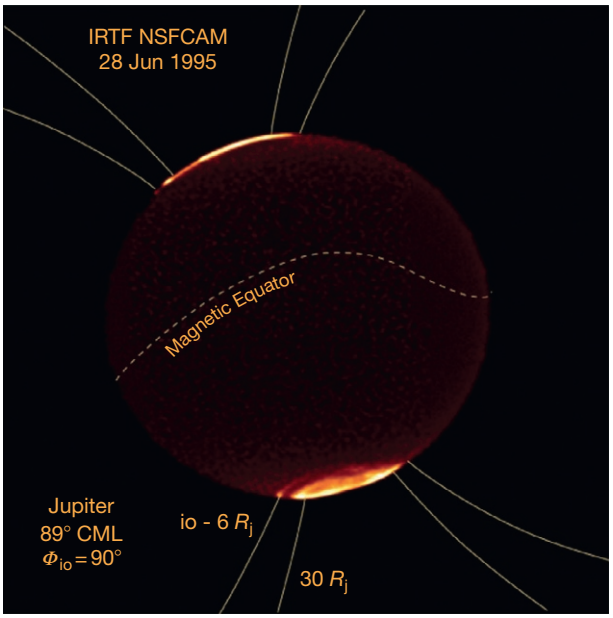


Figure 14 Image of Jupiter at 3.4 μm obtained at NASA's IRTF at Mauna Kea, Hawaii, using the NSFCAM facility imager. At this wavelength, bright H_3^+ auroral emissions originating above the homopause appear against a planetary disk darkened by methane absorption. Magnetic field lines are drawn in meridian plane projection to illustrate where field lines crossing the equator at $30R_j$ and $6R_j$ (Io's orbit) intersect Jupiter's surface. A bright emission feature at the foot of the Io flux tube (IFT) can be identified near the dawn limb, as well as faint emission extending along the Io L-shell downstream of the instantaneous IFT. Reproduced from Connerney JEP and Satoh T (2000) The H_3^+ ion: A remote diagnostic of the Jovian magnetosphere. *Philosophical Transactions of the Royal Society of London* 358: 2471–2483.

Jupiter's surface through which magnetic field lines with an equatorial crossing distance of $5.95R_j$ must pass. Thus, observations of the location (latitude, longitude) of the footprint offer a unique constraint on magnetic field models, precisely where it is most needed, on the surface of the planet, and otherwise unavailable. This emission is observable from Earth with ground telescopes (e.g., Infrared Telescope Facility on Mauna Kea) in the infrared region of the spectrum (Connerney and Satoh, 2000; Connerney et al., 1993) and with the Hubble Space Telescope in the ultraviolet region (Bonfond et al., 2009; Clarke et al., 1995, 1996, 2005; Prange et al., 1996). Emission has now been detected at the foot of the Europa and Ganymede flux tubes as well (Clarke et al., 2002; Grodent et al., 2006, 2009). The Europa and Ganymede observations are less useful constraints on the internal field because these field lines pass through the equator at greater radial distance (9.4 and $15R_j$, respectively) where relatively large and time-variable external fields are encountered. However, the Ganymede footprint supplies a useful fiducial marker for locating the source of the main aurora (Clarke et al., 2002), since emissions poleward of the Ganymede footprint must map to regions beyond its orbit ($15R_j$).

Over 100 determinations of the location of the IFT footprint were combined with *in situ* magnetic field observations (Pioneer 11 VHM and Voyager 1 FGM) to obtain a partial solution to a fourth-degree and fourth-order expansion of the internal field (Connerney et al., 1998). This model is referred to as the 'VIP4' model (**Table 3**), reflecting the use of Voyager 1, IFT, and Pioneer observations and an internal spherical harmonic expansion to degree and order 4. This model was designed to fit the position of the IFT footprint very well (within $\sim 1^\circ$ latitude), so it has proved particularly useful in analyses of satellite interactions and auroras. However, no one model fits the IFT footprints and *all* of the *in situ* magnetic field observations to within their expected errors (Connerney et al., 1998), although the IFT observations can be so fit with any *one* set of flyby observations.

Surprisingly, the IFT footprint observations are nearly sufficient *by themselves* to yield a useful field model. All that is needed is a means of constraining the magnitude of the field. The final model in **Table 3** lists a model derived using over 500 IFT footprint locations (Connerney et al., in preparation). This model constrains the field magnitude using a minimum of *in situ* magnetic field data: the theta component, only, of the Voyager 1 closest approach data obtained within $7R_j$. The Voyager 1 data were chosen as the most accurate ($\sim$0.1%) available and the theta component, ranging from 1036 nT to a maximum of 3280 nT, as the least influenced by external fields (larger in magnitude than the radial component of the external field but easily modeled). This model ('VIT4,' reflecting use of IFT observations, Voyager theta component, degree and order 4) is essentially a product of remote observation, using just enough *in situ* data to constrain the magnitude of the field. In principal, any constraint on the field magnitude – such as the field magnitude on the surface, deduced from the maximum frequency of radio emission at the electron cyclotron frequency – would serve as well if properly understood.

Spherical harmonic models confined to relatively low degree and order (e.g., 4) fit the IFT footprint observations well, but there remains a systematic and nonnegligible residual, particularly in the area of the 'kink' that appears in UV observations of satellite footprints and auroras near 110° Jovian system III longitude. Grodent et al. (2008) had suggested that a more localized source, or magnetic anomaly, might account for the 'kink.' This additional source could not be uniquely determined, but these authors demonstrated that the observations of satellite and auroral footprints in the northern hemisphere can be reproduced using a model of this type. The putative magnetic anomaly would necessarily have to be located near the surface in order to modify the satellite and auroral footprints without dramatically altering the field elsewhere. In perhaps the most ambitious such modeling attempt to date, Hess and coworkers utilized an extensive database of IFT footprints compiled by Bonfond et al. (2009) to obtain a fifth-degree and fifth-order spherical harmonic model ('VIPAL') of the Jovian magnetic field (Hess et al., 2011). This model obtains a better fit to the satellite footprints, at the expense of a poorer fit to the *in situ* magnetic field observations inward of about $4R_j$. However, VIPAL was also constrained to minimize the longitudinal separation between the observed and the modeled IFT footprints, and it was adjusted to produce a prescribed surface magnetic field strength along the IFT 'footpath.' This was done to better match the local electron gyrofrequency to observations of Jovian radio emissions, under certain assumptions regarding the source location, beaming, and generation of radio emissions. Thus, models of this type (e.g., VIPAL) require a certain degree of confidence in our understanding of the propagation of Alfven waves from Io to Jupiter (longitude constraint) and a degree of confidence in our understanding of the generation and propagation of Jovian radio emissions (surface magnitude constraint) as the price to be paid for encompassing more observables. In contrast, models based entirely on *in situ* magnetic field observations rest on our unshakable faith in the extrapolation of a potential field through a region of space devoid of local currents. In a few short years, the Juno spacecraft will enter Jupiter orbit (July 2016) to map Jupiter's magnetic field both closer to the planet and with much greater accuracy than ever before. Shortly after Juno's prime mission begins, it should be possible to discriminate among the approaches outlined earlier.

10.06.4.1.3 Discussion

The magnitude of the surface field for the O_6 model ranges from just over 3 G at low latitudes to just over 14 G at high (northern) latitudes. This is similar to the range of field magnitudes predicted by the O_4 model of Acuña and Ness (1976). The variation of magnetic field magnitude on the surface of Jupiter is illustrated in **Figure 15**, which depicts contours of constant field magnitude on the dynamically flattened surface of Jupiter, computed using the VIT4 model. This model has a less prominent north–south asymmetry in the magnetic field magnitudes, compared to earlier models; the maximum field in the northern hemisphere (13.9 G) is comparable to that of earlier models (see Acuña et al., 1983a,b), but the maximum field in the southern hemisphere (13.2 G) is considerably larger. The estimated uncertainty in the surface field magnitude is about ± 1 G (Connerney, 1981, 1992) but could be appreciably greater if higher-degree and higher-order harmonics are larger than expected or, equivalently, if a near-surface

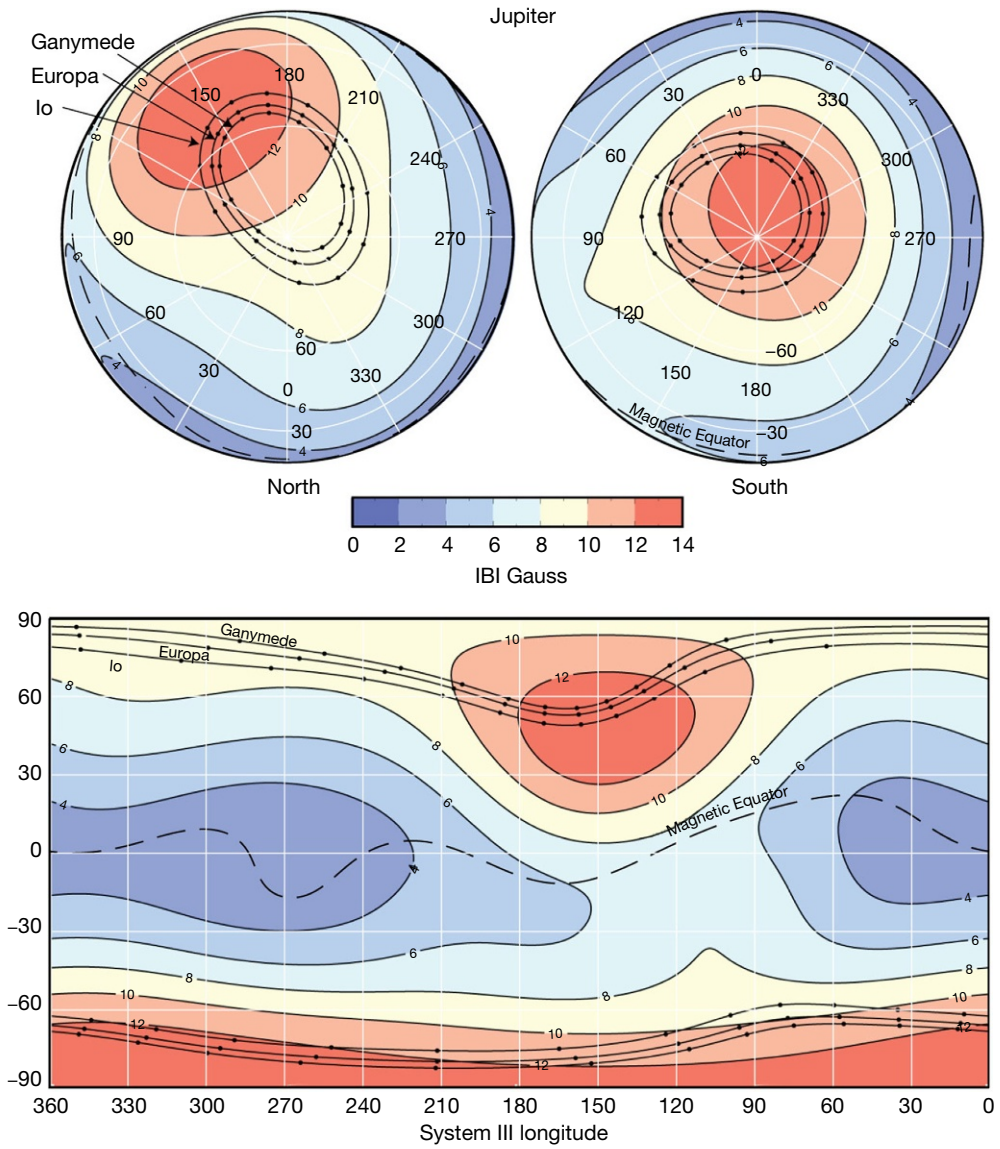


Figure 15 Contours of constant magnetic field magnitude (Gauss) on the dynamically flattened (1/15.4) surface of Jupiter, computed using the VIT4 model field (see text). The top panel shows the field magnitude in orthographic projections viewed from the north (left) and south (right); below the color bar is a rectangular latitude–longitude (system III west longitude) projection. A trace (dashed) indicates the position of the magnetic equator. Solid lines join points on the surface that trace along field lines to the orbits of the Galilean satellites Io, Europa, and Ganymede, with filled circles for increments of 30° in the satellite's system III longitude.

'magnetic anomaly' is invoked (e.g., Grodent et al., 2008). Modeled surface magnetic field magnitudes for the VIP4 and VIPAL models are not dramatically different (Hess et al., 2011); however, Grodent et al.'s magnetic anomaly model results in much larger surface field magnitudes in the northern mid-latitudes near 150°–180° system III longitudes. Figure 15 also illustrates the foot of the IFT ('Io foot'), which is the path traced out on the planet's surface by the field lines that pass through the satellite Io as the planet rotates. The foot of the IFT passes through the region of highest field strength in the northern hemisphere at a longitude of about 150° λ_{III}. The maximum surface magnetic field magnitude present along the Io foot is generally consistent with the maximum frequency extent (39.5 MHz) of Jovian decameter radio emission, assuming that the emission occurs at the local electron gyrofrequency and at the foot of the IFT (f_c(MHz) = 2.8B (G)). However, significant challenges arise in modeling Jovian radio emissions using such models, and this has motivated a number of authors to explore alternatives.

10.06.4.2 Saturn

Saturn, the second largest (R_s = 60 330 km) and most massive (M_j = 95M_e) planet in the solar system, resides at a mean orbital distance of 9.5 AU, nearly twice as far from the Sun as Jupiter. Like Jupiter, Saturn exhibits the characteristics of a condensed object of solar composition. Saturn was assumed to be much like its larger companion before Pioneer 11 arrived.

However, unlike Jupiter, Saturn as a radio source was relatively quiet. Saturn is too weak a radio source and too distant to be easily detected from Earth. In addition, Saturn emits at lower frequencies (a few kilohertz to 1 MHz) than Jupiter, well below the critical frequency required to penetrate the Earth's ionosphere. Thus, Saturn's radio emissions cannot be routinely monitored using terrestrial radio receivers.

10.06.4.2.1 Observations

The discovery of Saturn's magnetic field was thus left to Pioneer 11, arriving in September 1979, following its Jupiter swingby. The particle and field investigations on board Pioneer 11 (see previous section) obtained observations along a spacecraft trajectory that remained within $\sim 6^\circ$ of the equator (**Figure 16**). The most remarkable feature of Saturn's magnetic field proved to be the close alignment of the magnetic dipole and rotation axes (Acuña and Ness, 1980; Acuña et al., 1980; Smith et al., 1980a), also noted by the charged particle investigations (Simpson et al., 1980; Van Allen et al., 1980). Pioneer 11 approached to within $1.35R_s$, measuring a maximum field of ~8200 nT at -6° latitude (Smith et al., 1980a,b).

The Voyager 1 encounter in November 1980 and the Voyager 2 encounter in August 1981 provided the first observations of Saturn's magnetic field at high northern and southern latitudes (Ness et al., 1981, 1982). Voyager 1 sampled relatively high latitudes (-40°) at close radial distance ($3.07R_s$) but remained in the southern hemisphere while inside of $6R_s$ radial distance. Voyager 1 measured a maximum field of 1093 nT at -40° latitude and 184° SLS longitude, just prior to closest approach (Ness et al., 1981). Voyager 2's closest approach of $2.69R_s$ occurred at 323° longitude, diametrically opposite to those of Pioneer 11 and Voyager 1. Voyager 2 sampled relatively high latitudes in both hemispheres and measured a maximum field of 1187 nT just prior to closest approach (Ness et al., 1982).

More than two decades passed before the arrival of Cassini at Saturn. The Cassini spacecraft has been in orbit about Saturn since its orbit insertion (SOI) on 30 June 2004 and has remained relatively distant subsequent to the orbit insertion maneuver (Dougherty et al., 2004), targeting multiple satellite flybys from the vantage point of a near-equatorial orbit. During SOI, the spacecraft periapsis was just $1.33R_s$; this is the closest Cassini would come to Saturn throughout its prime mission. Cassini is instrumented with two magnetometers, a vector fluxgate and a helium magnetometer, the latter operated in a mode to measure field magnitudes near the planet (Dougherty et al., 2004). However, a large number of orbits have been executed thus far, many with periapsides $<4R_s$. The spacecraft is just now executing a series of maneuvers designed to increase the inclination of the orbit, which will provide a more favorable distribution of observations for internal field analysis, particularly near the end of mission, with many orbits planned with periapsides $<3R_s$.

10.06.4.2.2 Models

The Pioneer 11 magnetic field observations were consistent with a dipole field of moment ~0.20 G-R_s^3, slightly offset to the north of the planet's center (Acuña and Ness, 1980; Acuña et al., 1980; Smith et al., 1980a) by about 0.04 or $0.05R_s$. The simple northward-offset dipole model was affirmed by independent analyses of charged particle absorption signatures (Chenette and Davis, 1982), which serve as a probe of the geometry of the field. Analyses of the Pioneer 11 high-field FGM observations (Acuña et al., 1980) found no departure of the field from axisymmetry; however, in Saturn's relatively weak field, the resolution of the high-field fluxgate was quite limited. Working with higher-resolution VHM observations, Smith et al. (1980b) proposed a dipole tilt of $\sim 1^\circ$, although the orientation of the dipole could not be determined.

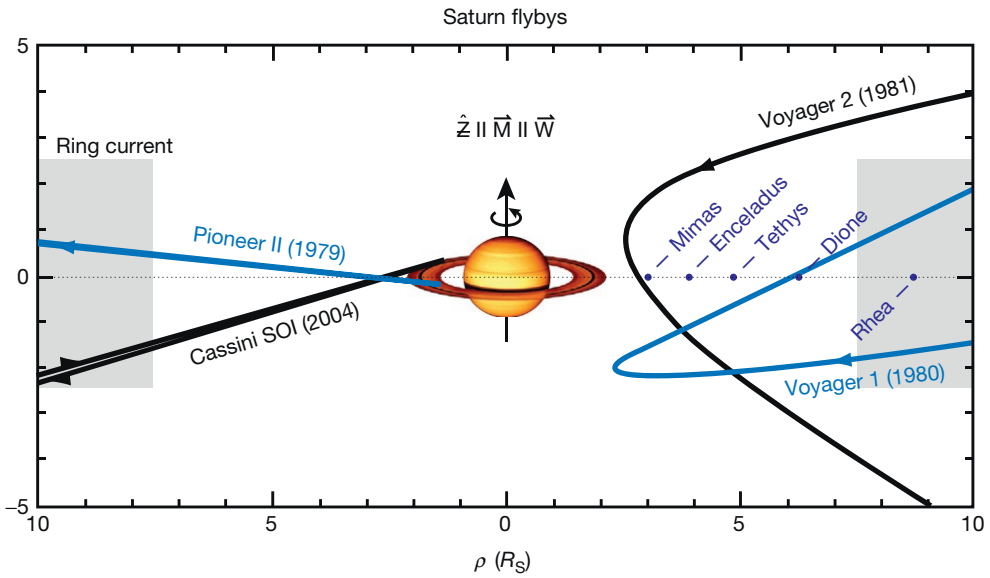


Figure 16 Spacecraft flybys (Pioneer 11, Voyager 1 and 2, and Cassini) of Saturn in a cylindrical coordinate system in which the z-axis is aligned with Saturn's magnetic dipole axis and rotation axis. The spacecraft distance from the equator plane (z) as a function of distance from the z-axis (ρ) is given in units of Saturn radius. The positions of satellites Mimas, Enceladus, Tethys, Dione, and Rhea are indicated. The shaded region approximates the innermost portion of a washer-shaped region of azimuthal ring currents.

The combined Voyager 1 and 2 magnetic field observations were not consistent with the field of a simple displaced dipole, nor did these observations reveal any measurable departure from spin axisymmetry (of the internal field). An axisymmetric, zonal harmonic model of degree 3 was found to be both necessary and sufficient to describe the magnetic field of Saturn (Connerney et al., 1982b). This three-parameter model is referred to as the Z_3 model (zonal harmonic of degree 3) and it is listed in **Table 4**; the nonaxisymmetric terms ($m>0$) are insignificant. Independent analyses of the Voyager 1 and 2 observations suggested an estimated uncertainty of ~100 nT for the Z_3 coefficients. This is perhaps the best way to estimate errors when model residuals are highly correlated – a comparison of independent analyses, using separate flybys. The Z_3 model parameter uncertainties are likely to be smaller, benefiting from the combined observations of both Voyager flybys. The magnetic and rotation axes are indistinguishable, aligned to within ~0.1°.

A series of analyses followed to test the high degree of axisymmetry of the Z_3 model (reviewed in Connerney et al. (1984b)). Acuña et al. (1983a) showed that the charged particle absorption signatures observed in Saturn's magnetosphere were consistent with the Z_3 model. A reanalysis of the absorption signatures studied by Chenette and Davis (1982), allowing for the presence of a nonzero g_3^0 term, found these signatures to be best fit by a zonal harmonic model that was indistinguishable from the Z_3 model (Connerney et al., 1984b). A reanalysis of the Pioneer 11 VHM observations (Connerney et al., 1984a) identified a spacecraft roll attitude error (of 1.4° in magnitude) as the primary source of the discrepancy between the Pioneer 11 and the Voyager results. After correcting the spacecraft roll attitude, these authors demonstrated that the Pioneer 11 VHM measurements were consistent with the Z_3 model field, and axisymmetry, to better than the accuracy of measurement (Connerney et al., 1984a). The roll attitude error was subsequently independently verified by the analysis of Pioneer 11 imaging photopolarimeter data. After correcting for the Pioneer 11 roll attitude error, Davis and Smith (1985, 1986) combined the Pioneer 11 VHM and Voyager magnetometer observations and obtained a zonal harmonic model Saturn Pioneer Voyager (SPV) in substantial agreement with the Z_3 model (**Table 4**).

Table 4 Saturn magnetic field models

	Cassini	*Voyager*			*Pioneer 11, Voyagers 1 and 2*
Epoch	*2004.5*	*(V1+V2)*	*1980.9*	*1981.6*	*1979–1981*
Model	*SOI*	*Z_3*	*V1*	*V2*	*SPV*
Zonal harmonics					
g_1^0	21084.0	21184.0	21224.0	21084.0	21160.0
g_2^0	1544.0	1606.0	1678.0	1607.0	1560.0
g_3^0	2150.0	2669.0	2613.0	2513.0	2320.0
External field					
G_1^0	( )	−10	−11	−8	( )
G_1^1		−1	0	−1	( )
H_1^1		0	0	−2	( )

Schmidt normalized zonal harmonic coefficients in nT, SLS coordinate system: $1R_s = 60\,330$ km for Voyager and Pioneer 11 models. Cassini SOI uses $1R_s = 60\,268$ km. Voyager and Pioneer models provided in this table have been adjusted for a reference radius of 60 330 km; see Connerney et al., 1982b for coefficients referenced to 60 000 km radius as originally published. The SPV model (Davis and Smith, 1990) used Pioneer 11 observations corrected for a roll attitude error (Connerney et al., 1984a) in combination with Voyager observations. SPV used different external (uniform) fields for each encounter.

Analysis of the Cassini FGM observations obtained during SOI confirmed (to the level of estimated measurement accuracy, ~0.1% and 0.1°) the spin symmetry of Saturn's magnetic field (Dougherty et al., 2005). An unfortunate command error prevented the helium magnetometer from making measurements of the field magnitude during SOI. A zonal harmonic model of degree 3 was obtained (**Table 4**) from the fluxgate observations; it is very similar to the earlier zonal harmonic models derived from Voyager and Pioneer observations but for a smaller g_3^0 term. **Figure 17**, in which we plot the vector magnetic field observed throughout Cassini's orbit insertion maneuver, illustrates the remarkable spin symmetry of Saturn's internal field. The absence of a significant azimuthal field component during this close encounter suggests a negligible tilt of the dipole and the lack of higher-order contributions to the field.

More revealing, perhaps, is **Figure 18**, which plots the vector residual after the best-fitting zonal harmonic model has been removed from the observations. Residuals that *are* observed can be understood by attention to engineering practicalities, for example, spacecraft attitude maneuvers associated with turns to avoid impact of ring material on sensitive components ('turn turtle,' i.e., assume a protected attitude crossing the ring plane) and the Earth point maneuver. Errors in the measured vector magnetic field can arise from attitude uncertainties associated with spacecraft maneuvers or with variations in sensor response as the field direction in the sensor reference frame changes. A turning spacecraft often depends on gyros for attitude information, lacking a celestial reference, and slight errors in gyro drift rates may result in attitude errors during maneuvers. Likewise, one sees an increase in variations of the field associated with the main engine burn and its aftermath. This figure also illustrates that the residuals are highly correlated and dominated by systematic effects, most of which can be readily understood and related to spacecraft engineering events; thus, estimates of model parameter uncertainties based on propagation of errors and Gaussian statistics are likely to be overly optimistic.

Near the end of Cassini's prime mission, a number of orbits were acquired with periapsides inside the orbit of Enceladus ($3.95R_s$), providing much-needed close-in observations at higher latitudes (though still within about ±30° of the equator). More recent modeling efforts have assembled a large number of orbits obtained over a lengthy span of time, fitting zonal harmonic internal fields with external fields described by external spherical harmonics or an explicit ring current (or magnetodisk) model. These models are listed in **Table 5**. Burton et al. (2009) analyzed observations (45 individual orbits) acquired during Cassini's first 3 years (July 2004–2007) utilizing an explicit ring current model to represent external fields. They found no departure from axisymmetry in the internal field and little evidence for secular variation of the field over the 30 years between the Pioneer/Voyager era and the Cassini era. A later study by the same authors (Burton et al., 2010) used observations confined to the region inside the orbit

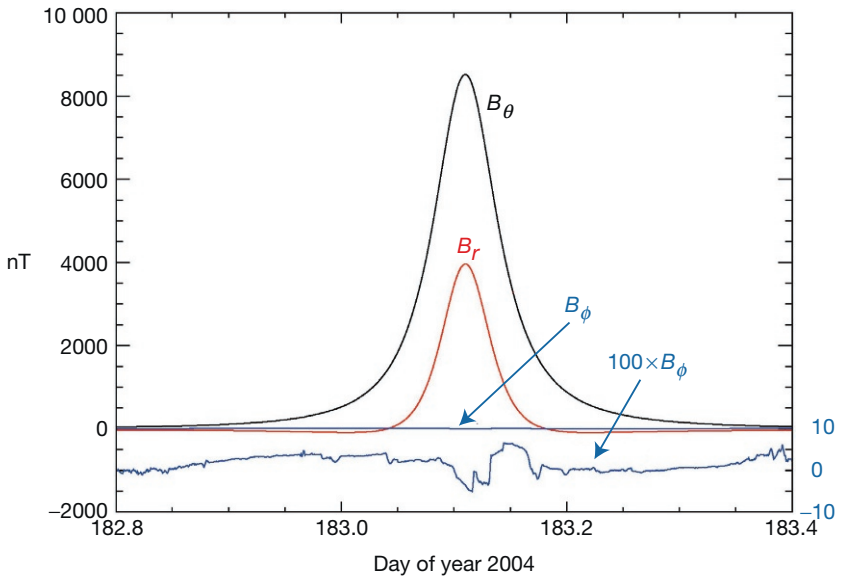


Figure 17 Vector magnetic field observed during Cassini Saturn Orbit Insertion (SOI) on day 183, 2004. The observed field is dominated by the smooth variation of the radial (red) and theta (black) components of the field. The azimuthal component of the magnetic field (blue) is very small throughout the equatorial passage and is plotted again increased by a factor of 100 and offset by 10 nT for clarity.

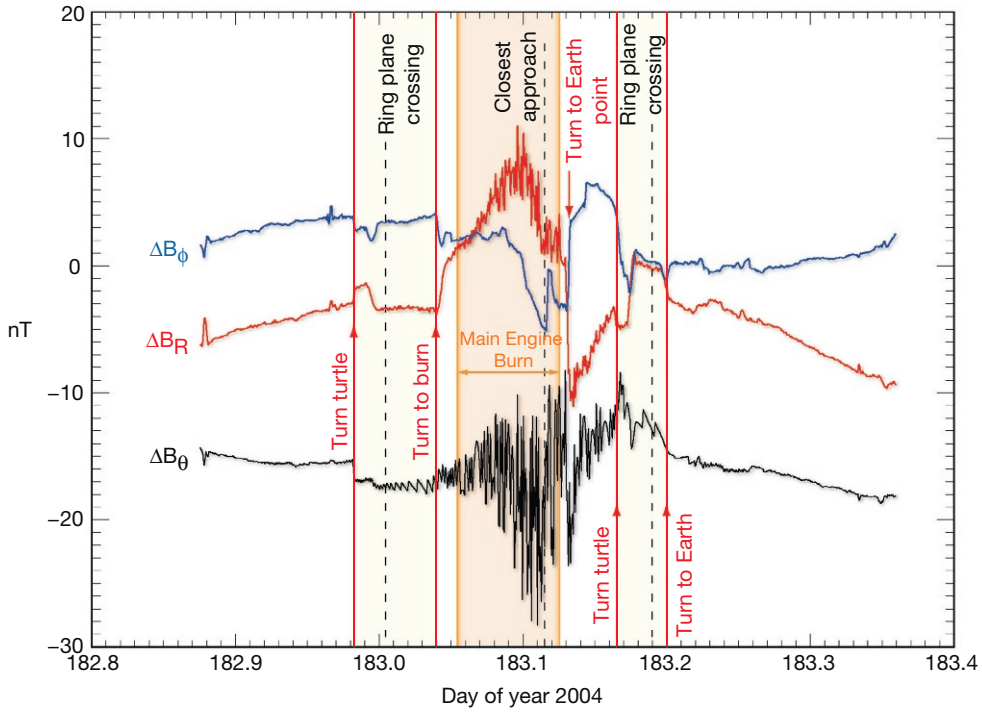


Figure 18 Vector residual magnetic field during the Cassini Saturn Orbit Insertion on day 183, 2004. The best-fitting zonal harmonic (axisymmetric) model of Saturn's planetary magnetic field has been removed from the observations, and the components of the residual field have been plotted as a function of time. The radial and theta components are offset (by 10 and 20 nT, respectively) for clarity. Significant spacecraft maneuvers and events are indicated. Systematic differences between the best-fitting zonal harmonic model and the observed field are evident; the residuals in all three components are clearly associated with spacecraft turns and propulsive events.

of Enceladus at $3.95R_s$ (to mitigate external effects) and represented external fields with a spherical harmonic expansion, reaching the same conclusions listed earlier. The latter study also explored a range of putative Saturn rotation periods, finding no preference for a different rotation rate and no significant nonaxial internal field. Likewise, Cao et al. (2011) modeled observations obtained over a span of 5 years within the Enceladus L-shell, using a simple uniform external field

representation. They also found that a zonal harmonic internal field of degree three is sufficient to describe Saturn's magnetic field. They arrived at a reduced upper limit for the tilt of the dipole ($<0.06°$). Their more recent reanalysis of the Cassini SOI observations (Cao et al., 2012) – fixing the zonal harmonics with degrees 1–3 and performing a Monte Carlo search for a best-fitting g_4^0 and g_5^0 – resulted in the identification of nonzero fourth- and fifth-degree zonal harmonic terms. These are both very small in magnitude, however, and not much larger than the estimated parameter uncertainties (which are likely to be quite optimistic; see, e.g., **Figure 18**).

Cassini, in its extended 'Solstice' mission, now enters a phase in which the orbital inclination increases, providing close-in observations of the magnetic field at progressively higher latitudes. With better spatial coverage, and as more observations accumulate, one may expect some refinement of the spherical harmonic models derived from Cassini.

10.06.4.2.3 Discussion

Magnetic field magnitudes (Z_3 model) on the surface of a dynamically flattened Saturn (flattening 1/10.6) range from a minimum of 0.18 G at the equator to maxima of 0.65 and 0.84 G at the southern and northern poles, respectively (**Figure 19**). Note that Saturn is the most dynamically flattened of all the planets, and the polar magnetic field is substantially increased by the decreased radial distance ($0.9R_s$) at the poles. The equatorial field magnitude is less than that of a simple dipole or offset dipole, and polar field magnitudes (both north and south) exceed those of either a centered or an offset dipole. An eccentric dipole representation of the Z_3 model has a northward displacement of the dipole of $0.038R_s$ along the rotation axis (Connerney, 1993), in agreement with the Pioneer 11 OTD representations. However, the ratio of g_3^0/g_2^0 is too large by a factor of 30 to be attributed to a dipole offset. The Z_3 model's large antisymmetry with respect to the equator plane cannot be removed by a simple coordinate system translation. The same can be said of all of the Saturn zonal harmonic models listed in **Tables 4** and **5**.

The high degree of spin axisymmetry of Saturn's magnetic field is unique among the planets. This unanticipated symmetry is somewhat disquieting from an observational perspective, as it lends no simple explanation for observations of periodic magnetospheric phenomena, for example, periodicities in magnetic field and particle observations in the distant magnetosphere (Connerney and Desch, 1992; Cowley et al., 2006; Galopeau et al., 1991; Giampieri and Dougherty, 2004; Giampieri et al., 2006; see also Kaiser et al., 2005 and Giampieri et al., 2005), the modulation of Saturn kilometric radio emission (Desch and Kaiser, 1981; Kaiser and Desch, 1982), optical spoke activity in Saturn's B ring (Porco and Danielson, 1982), and ultraviolet auroras (Sandel et al., 1982). All of these phenomena would presumably be easier to explain if a magnetic anomaly could be found. However, it is not clear how such a slight departure (e.g., near the $\sim 0.1°$ limit of detection) from axisymmetry might so

Table 5 Saturn Cassini orbiter magnetic field models

	Cao et al. (2011)	*Burton et al. (2010)*	*Burton et al. (2009)*
Epoch	*2005–2010*	*2004–2008*	*2004–2007*
Zonal harmonics			
g_1^0	21191.0 (±24)	21136.0	21162.0
g_2^0	1586.0 (±7)	1526.0	1514.0
g_3^0	2374.0 (±47)	2219.0	2283.0
Tilt	$<0.06°$	$<0.1°$	

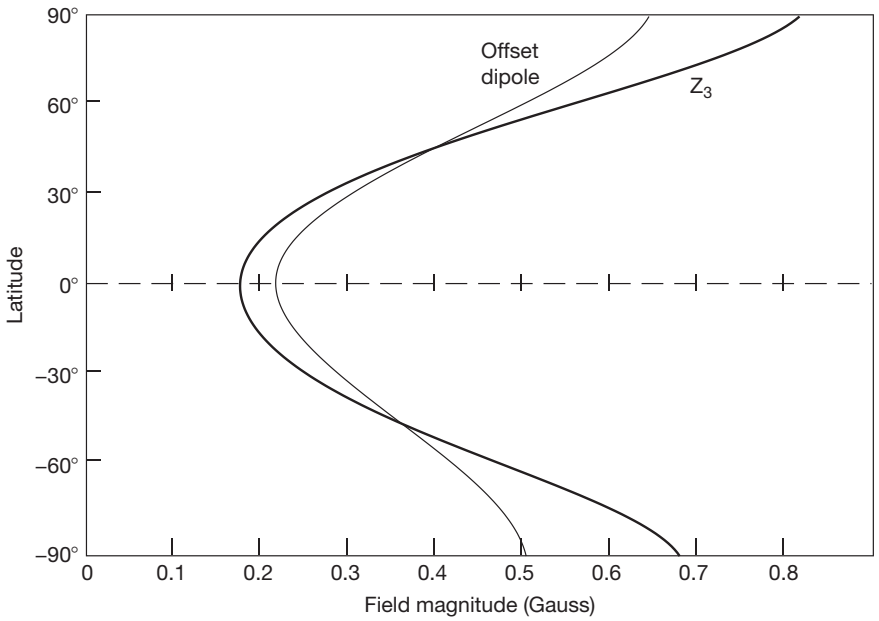


Figure 19 Magnetic field magnitude (Gauss) as a function of planetocentric latitude on the dynamically flattened (1/10.6) surface of Saturn computed using the axisymmetric Z_3 (zonal harmonic of degree 3) model of Saturn's magnetic field (Connerney et al., 1982b), compared with that of a simple offset dipole.

dramatically influence these phenomena, which occur at different magnetic latitudes and as far as 1 or $2R_s$ above the surface.

The uncertainty regarding the origin of Saturn's quasi-periodic magnetospheric phenomena has spilled over into a related topic: Saturn's rotation period. Accurate measurement of the rotation period of gaseous planets relies on the long-term observation of magnetospheric radio emissions, which are assumed to be locked in phase with the magnetic field and interior of the planet. Long-term observation of Jupiter's radio emissions, for example, yield a consistent estimate (Riddle and Warwick, 1976) of rotation period (9 h, 55 m, 29.71 s) regardless of frequency of emission measured or method of analysis (Carr et al., 1983). Note that the uncertainty associated with this estimate is about 0.02 s (Carr et al., 1983), reflecting observation over many years. Saturn's rotation period was measured in the Voyager era (1980–81) using periodicities in kilometric radio emissions (SKR), yielding an estimate of the rotation period (10 h, 39 m, 24 s) with an uncertainty of ± 7 s (Desch and Kaiser, 1981; Kaiser et al., 1984). The larger uncertainty of this estimate is due to the limited span of time (9 months) over which SKR could be monitored. Continued observation of Saturn's radio emissions, however, in the Ulysses era (Gurnett et al., 2005) and during the Cassini mission (Gurnett et al., 2006, 2007; Kurth et al., 2007) revealed variations of about $\pm 1\%$ ($\sim$6 min) in the apparent radio period – far in excess of that allowed by statistical uncertainties. Likewise, direct measurement of periodicities in the magnetic field (Giampieri and Dougherty, 2004; see also Cowley et al., 2006; Giampieri et al., 2005, 2006; Kaiser et al., 2005) yielded inconsistent rotation periods, at best, with little promise of establishing a true rotation rate (Sterenborg and Bloxham, 2010).

Saturn's radio emissions are influenced by variations in the solar wind (Desch, 1982; Desch and Rucker, 1983) and are emitted from a source region fixed in local time (Galopeau et al., 1995; Kaiser and Desch, 1982; Kaiser et al., 1984). It has thus been proposed that variations in the apparent periodicity of radio emissions are related to variations in the solar wind (Cecconi and Zarka, 2005). Gurnett et al. (2007) observed that the SKR period derived from Cassini radio observations differed substantially from that deduced from the Voyagers; even worse, the SKR-derived period was found to be anything but constant in time (Gurnett et al., 2007). Thus, it is not clear which of the proposed rotation periods, if any, closely match that of the deep interior of the planet.

Not surprisingly, then, interest has turned to other means of establishing Saturn's rotation period and even questioning what it means to assign such a number to a fluid planet (Stevenson, 2006). The Voyager era rotation period was never warmly embraced among atmospheric scientists because this period results in an atmosphere that superrotates at nearly all latitudes (Melendo et al., 2011; Smith et al., 1982). In contrast, Jupiter's atmosphere is characterized by alternating eastward and westward jets, referenced to its system III (radio) period. Thus, a number of proposals have emerged that would have Saturn rotating significantly faster. Anderson and Schubert (2007) treated Saturn's rotation period as a free variable and found the period (10 h, 32 m, 35 ± 13 s) that best reconciles gravity, radio occultation, and wind observations; they proposed this as a more accurate period of rotation of Saturn's interior. The validity of this method is somewhat bolstered by application to Jupiter (Helled et al., 2009), where it results in a period that agrees well with the radio-derived period. A similar Saturn rotation period (10 h, 34 m, 13 ± 20 s) is obtained from a consideration of atmospheric dynamic stability (Read et al., 2009; see also Showman, 2009). Either proposal results in a more Jupiter-like system of eastward- and westward-flowing jets that would be expected of a giant planet with a water-rich atmosphere driven by large-scale latent heating (Lian and Showman, 2010).

This is the subject of intense activity at present, so perhaps, a better understanding of Saturn's rotation will be forthcoming soon. An accurate value for the rotation period is essential to arriving at an understanding of Saturn's interior structure and composition.

The magnetic fields of Earth, Jupiter, and Mercury all favor dipole tilts of $\sim 10^\circ$ in magnitude; Saturn's dipole tilt is not yet determined but $<0.06^\circ$ or so. The high degree of spin axisymmetry is surprising but not in and of itself a problem for dynamo theory (e.g., Lortz, 1972). The Cowling antidynamo theorem (Cowling, 1937, 1957), which states that regenerative dynamo action cannot occur if the magnetic field and fluid motions are axisymmetric, applies to the field in the dynamo region, not the field observed above. It is, however, necessary to explain why Saturn's magnetic field is so dramatically different from Jupiter's, given the similarity in size, composition, and rotation.

Stevenson proposed a model in which the field is axisymmetrized by the differential rotation of an electrically conducting shell above the dynamo region (Stevenson, 1980, 1982; also Kirk and Stevenson, 1987). This nonconvecting shell forms above the dynamo region as a result of the immiscibility of helium and hydrogen under the temperature and pressure conditions in this region (the larger planet, Jupiter, would have no such region). In this model, the formation of helium raindrops results in a stably stratified conducting shell which, in differential rotation with respect to the core, attenuates any nonaxisymmetric fields generated below. Schubert et al. (2004) treated the more general case of dynamo generation in a convective core within an electrically conducting shell, taking into account the mutual coupling between the two. These authors concluded that the observable characteristics of planetary dynamos are largely determined by the nature of the (electrically conducting) flow above the dynamo region. That is to say, the mere presence of a stably stratified, conducting shell in differential rotation above the dynamo is not sufficient to axisymmetrize the field (Stanley and Mohammadi, 2008); but it may act to axisymmetrize the field if it is sustained by a thermal wind as a result of a pole-to-equator temperature difference (Christensen and Wicht, 2008; Stanley, 2010). The dynamo models studied thus far do not go quite far enough in axisymmetrizing the field, producing dipoles with 1–2° angular separation between magnetic and spin axes, compared to Saturn's $<0.1^\circ$ dipole tilt.

The Z_3 model of Saturn's field can be used (with caution) to predict the field inside the planet but above the dynamo (or regions of current flow). The radial field component is of particular interest since it is continuous across the core boundary and, if the frozen flux assumption is satisfied, serves as a tracer of horizontal fluid motion at the outer radius of the dynamo (Backus, 1968; Benton, 1979; Benton and Muth, 1979). **Figure 20** shows the latitudinal variation of $|B_r|$,

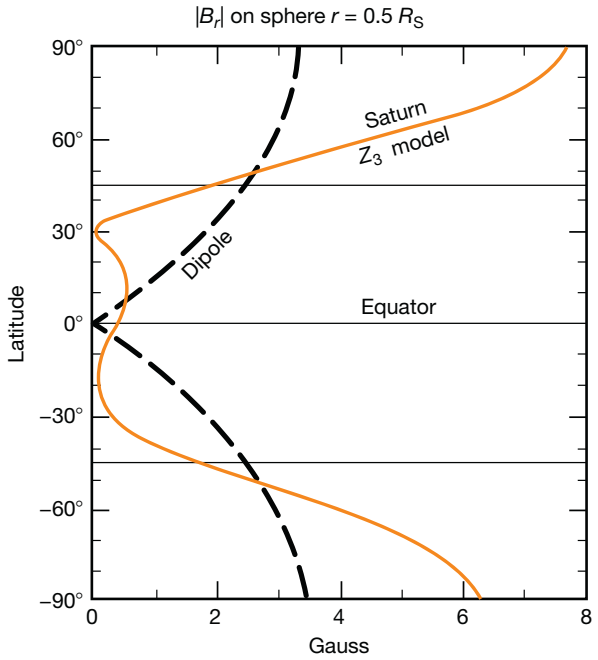


Figure 20 Unsigned magnetic flux computed on a sphere of radius 0.5 R_s (presumed dynamo core radius) using the Z_3 magnetic field model, compared to that for a simple dipole. Saturn's magnetic field is very nearly identical to that that (with three terms) minimizes the low-latitude (<45°) unsigned magnetic flux across the core boundary. Reproduced from Connerney JEP (1993). Magnetic fields of the outer planets. *Journal of Geophysical Research* 98(E10): 18659–18679.

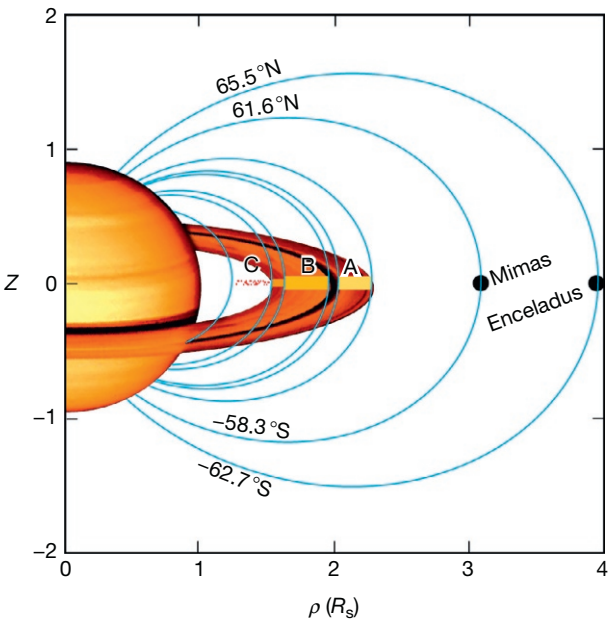


Figure 21 Saturn's magnetic field is very nearly symmetrical about its rotation axis but not about the equator. All field lines passing through the ionosphere at one latitude cross the ring plane at a unique radial distance (ring plane conjugate) and map to a corresponding (but different) latitude in the opposite hemisphere. The magnetic equator is at 5.9° N (planetocentric) latitude. Reproduced from Connerney JEP (1986) Magnetic connection for Saturn's rings and atmosphere. *Journal of Geophysical Research* 13(8): 773–776.

computed using the Z_3 model, on the surface of a sphere of radius $r = 0.5R_s$, which is the approximate outer boundary of the metallically conducting core (Stevenson, 1983). The Z_3 model (via the g_3^0 term) enhances the field at the poles and reduces the radial field near the equator, relative to a simple dipole. In fact, the Z_3 model is very nearly identical to that required to minimize the low-latitude unsigned flux across the core surface (Connerney, 1993). Cao et al. (2012) reached a similar conclusion with their preferred zonal harmonic model, albeit with a dynamo core radius ($0.4R_s$) that is significantly smaller than the radius ($0.64\ R_s$) of metallically conductive hydrogen (Weir et al., 1996). The concentration of radial flux at the spin poles, and minimization of radial flux at low latitudes, may be the signature of a dynamo axisymmetrized by differential rotation of a conductive layer above.

The high degree of symmetry (about the rotation axis) of Saturn's magnetic field is a distinguishing feature that may have some unexpected consequences throughout the system. For this planet, a specific radial distance in the equatorial plane maps (along the magnetic field) uniquely to a corresponding latitude on the surface (Connerney, 1986), as is illustrated in Figure 21. Thus, one may associate a radial distance (ring plane conjugate) with any northern or southern latitude. From a practical standpoint, all of the zonal harmonic models introduced subsequent to the Z_3 model yield a very similar mapping from equator to ionosphere. Narrow dark bands that encircle the planet at 44.2°, 46.3°, and 65.5° N have been attributed (Connerney, 1986) to an influx of water guided along magnetic field lines from sources in the ring plane: two locations at the inner edge of the B ring ($1.525R_s$, $1.62R_s$) where high-charge-to-mass-ratio particles (e.g., submicron ice or dust) become unstable (Northrop and Connerney, 1987) and a third ($3.95R_s$) at the orbit of Enceladus and the E ring. The B-ring sources are identified with an electromagnetic erosion mechanism that sculpted the rings (Northrop and Connerney, 1987), a process that over tens of millions of years removed most of the mass from the C ring, for example. Recent Keck observations of H_3^+ emissions from Saturn's ionosphere (O'Donoghue et al., 2013) reveal a latitudinal variation in ion density that suggests transport of water in the form of high-charge-to-mass-ratio particles along magnetic field lines from sources in the rings (Connerney, 2013). The Enceladus/E-ring source may reflect persistent activity of the recently discovered geysers on Enceladus (Kargel, 2006; Kivelson, 2006).

10.06.5 Ice Giants

10.06.5.1 Uranus

Uranus, residing at a mean orbital distance of 19.2 AU, was a mystery prior to the Voyager 2 encounter with Uranus in January 1986. The planet appears bland and featureless when observed from the Earth even under the best circumstances. Uranus and Neptune are similar in size (~25 000 km radius), mass (about 15 Earth masses), and presumably composition (largely low-temperature condensates or 'ice': H_2O, NH_3, and CH_4). One might therefore expect that their magnetic fields would be similar. Uranus is unique in that its rotation axis lies

very near in its orbital plane, which is within 1° of the ecliptic plane. The actual inclination of its equatorial plane to that of its orbit is 82°. Twice during each orbit about the Sun (84 years), the rotation axis is oriented very near along the planet–Sun line, so that we on Earth look upon the pole. This was the case during the Voyager 2 encounter in 1986; the southern hemisphere was directed toward the Sun.

Uranus was not known to have a magnetic field prior to the Voyager encounter, having maintained 'radio silence' prior to encounter. It is a relatively weak radio source, emitting only at lower frequencies (below about 1 MHz), making detection from Earth difficult. In addition, the peculiar 'pole-on' encounter geometry kept the Voyager spacecraft beyond the reach of radio emissions. Voyager 2, on a flyby trajectory (**Figure 22**) approaching to within $4.2R_u$, provided the first and thus far only measurements of the magnetic field (Ness et al., 1986). These measurements led to the rather surprising discovery of the first 'oblique rotator,' a planet with a large angular separation between magnetic and rotation axes.

10.06.5.1.1 Models

Preliminary analyses of the Voyager magnetic field observations used an OTD representation of the field. This model is characterized by a dipole of moment 0.23 G R_u^3 ($1R_u = 25\,600$ km), displaced along the rotation axis by $0.33R_u$ and inclined by 60° with respect to the rotation axis (Ness et al., 1986). Since Voyager remained somewhat distant, an OTD representation is expected to be a good approximation to the field, and indeed, it fits the observations well (Ness et al., 1986). The very large dipole offset of the OTD is an indication of a planetary magnetic field with a large quadrupole and likely higher-degree and higher-order terms, that is, a very complex field.

A spherical harmonic analysis of the field required an expansion to degree and order 3 (octupole) to adequately represent the measurements (Connerney et al., 1987), which is somewhat surprising in view of Voyager's relatively large periapsis ($4.2R_u$). The Voyager trajectory provided a fairly good spatial distribution of observations, sufficient to resolve, or determine, spherical harmonic terms of degrees 1 and 2, the dipole and quadrupole. This partial solution to the underdetermined inverse problem, designated 'Q_3' (quadrupole part of an expansion to degree and order 3), is listed in **Table 6**. The relative magnitude of the quadrupole is quite large; the octupole is likely to be large as well, but cannot be meaningfully constrained by the observations alone. Holme and Bloxham (1996) obtained octupole coefficients utilizing a regularization technique that imposes a smoothness constraint on the field; the resulting model is quite similar to that obtained by Connerney et al. (1987) among the lower-degree coefficients and presents a plausible extension of the model field to higher degree and order. Lack of knowledge of higher-degree and higher-order terms may be an important consideration, particularly in estimation of the field near the planet. The Q_3 model was extended to higher degree and order using the mapped location of near-surface ultraviolet auroral emissions that were assumed to lie on conjugate L-shells (Herbert, 2009). The Q_3 model provides a low-degree and low-order approximation only, and it is likely that surface fields are more complex than illustrated here.

The magnetic field magnitude on the surface of Uranus reaches a maximum of ~1 G at midlatitudes in the south and

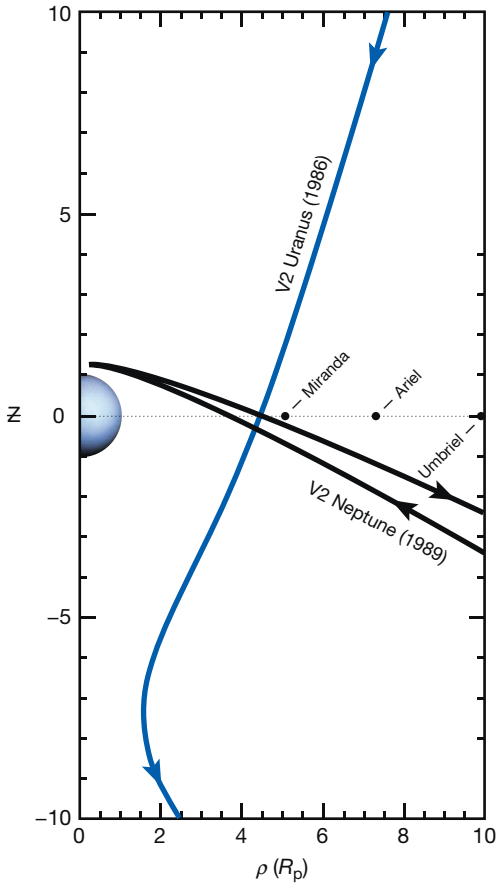


Figure 22 Voyager 2 spacecraft flybys of Uranus and Neptune in a cylindrical coordinate system. The spacecraft distance from the equator plane (z) as a function of distance from the rotation axis (ρ) is given in units of planet radius. The positions of satellites Miranda, Ariel, and Umbriel in the Uranus system are indicated.

Table 6 Uranus magnetic field models

		Connerney et al. (1987)	*Holme and Bloxham (1996)*
Coefficient		*Uranus* Q_3	*Uranus*
1	g_1^0	11893.0	11855.0
2	g_1^1	11579.0	11507.0
3	h_1^1	−15689.0	−15812.0
4	g_2^0	−6030.0	−5877.0
5	g_2^1	−12587.0	−13085.0
6	g_2^2	196.0	−605.0
7	h_2^1	6116.0	5851.0
8	h_2^2	4759.0	4185.0
9	g_3^0		4183.0
10	g_3^1		−1336.0
11	g_3^2		−6776.0
12	g_3^3		−4021.0
13	h_3^1		−5817.0
14	h_3^2		−357.0
15	h_3^3		−2265.0

Spherical harmonic coefficients in nT; $1R_u = 25\,600$ km.

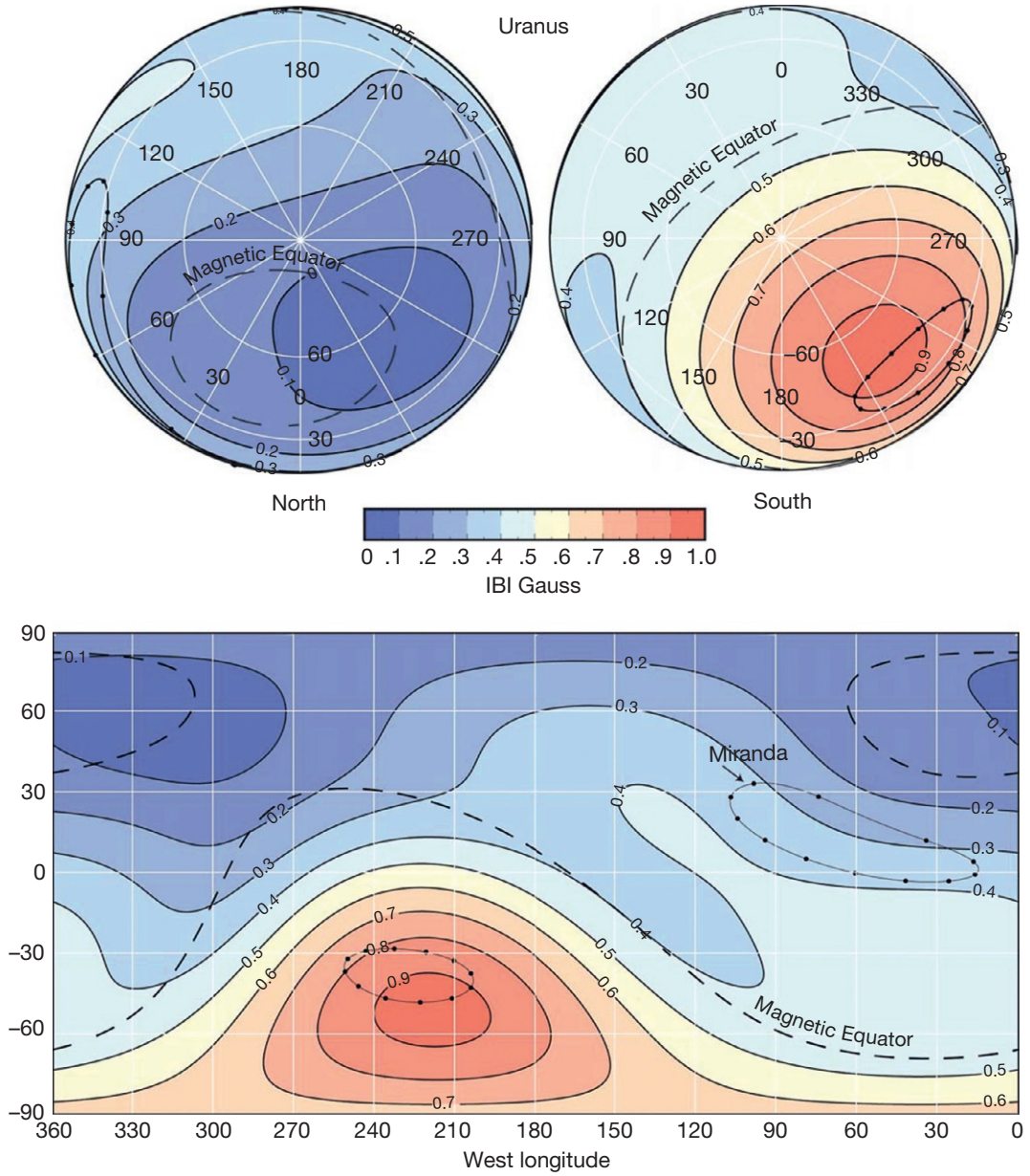


Figure 23 Contours of constant magnetic field magnitude (Gauss) on the dynamically flattened (1/41.6) surface of Uranus, computed using the Q_3 spherical harmonic model (see text). The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) to $-z$ (right); below the color bar is a rectangular latitude–longitude (west longitude) projection. A trace (dashed) indicates the position of the magnetic equator. Solid lines join points on the surface that trace along field lines to the orbit of the satellite Miranda, with filled circles for increments of 30° in the satellite's longitude.

a minimum of ≤ 0.1 G at middle to high northern latitudes (Figure 23). Figure 23 also illustrates the foot of the Miranda flux tube, which is the path traced out on the planet's surface by the field lines passing through the satellite as the planet rotates. Also shown are estimates of the position of the auroras (stippled regions) and the paths of multiple-dip equators (where $B_r = 0$). Uncertainties in the higher-degree and higher-order components of the field, which are substantial, may lead to large differences in the actual surface field, particularly in the weaker field regions. Note that for Uranus, the convention adopted by the Voyager particles and fields investigations is at odds with the International Astronomical Union (IAU) latitude convention. The Voyager system has planetary latitudes positive over the hemisphere in which the angular momentum vector resides, that is, the hemisphere oriented toward the Sun (and Earth) at the time of the Voyager encounter. The IAU defines the south pole as the rotation pole south of the ecliptic, without regard to the direction of rotation.

10.06.5.2 Neptune

Neptune, residing at a mean orbital distance of 30 AU, is the most distant of the Jovian planets. It is similar in size (~25 000 km radius), mass (about 15 Earth masses), and

composition (largely low-temperature condensates or 'ice': H_2O, NH_3, and CH_4) to Uranus. Very little was known of Neptune prior to the Voyager 2 encounter in August 1989. The first evidence of Neptune's magnetic field appeared in radio observations (Warwick et al., 1989) obtained by Voyager 2 just days prior to its closest approach to the planet on 25 August 1989.

The *in situ* magnetic field observations revealed a magnetic field of internal origin with characteristics not unlike those of the planet Uranus (Ness et al., 1989). The most striking difference between the Uranus and the Neptune encounters was a difference of our own design: the Voyager 2 flyby trajectories at each planet. Where Voyager 2 at Uranus remained relatively distant from the planet, Voyager 2 at Neptune approached to within $1.18R_n$ of the planet (**Figure 22**). The presence of higher-degree and higher-order components of the field in the latter observations is overwhelming, simply a consequence of the close periapsis. Preliminary analyses of the Voyager magnetic field observations used an OTD representation of the field, valid at moderate distances from the planet (4–15R_n radial distance; $1R_n = 24\,765$ km), and fitted to data in excess of $4R_n$ from the planet only. In this first approximation, the field may be characterized as that of a dipole of moment 0.13 G R_n^3, offset from the center of the planet by a rather surprisingly large $0.55R_n$ and inclined by 47° with respect to the rotation axis (Ness et al., 1989).

10.06.5.2.1 Models

It was necessary for Ness et al. (1989) to exclude from their OTD analysis all observations of the field obtained at $r \leq 4R_n$. One indication of how nondipolar the field of Neptune is can be had by comparison of the actual field measured at closest approach (~10 000 nT) to that one would expect if the field were dipolar (6500 nT, for the OTD model). In fact, the field measured near closest approach was quite complex, as is illustrated in **Figure 24**. Two local maxima appear in the magnitude of the field observed along the flyby trajectory, instead of the one maximum one would observe in a dipolar field. Spherical harmonic analysis of the field required an expansion to degree and order 8 to adequately represent the measurements (Connerney et al., 1991). The large number of parameters associated with a spherical harmonic expansion of degree and order 8 (80 Schmidt coefficients) and the rather limited observations lead to a severely underdetermined inverse problem that has only a partial solution (Connerney et al., 1991). The distribution of observations is sufficient to resolve, or determine, all of the spherical harmonic terms of degree 1 (dipole), some of the terms of degree 2 (quadrupole), and some terms of degree 3 (octupole) and higher. The resulting model, designated 'O_8' (octupole part of an expansion to degree and order 8), is listed in **Table 7**. For the Neptune field model, it is necessary to add a measure of the resolution of each parameter listed in **Table 7**. For this, we use the magnitude of the corresponding diagonal element of the resolution matrix, a measure of the uniqueness of the solution, the extent to which an individual model parameter can be represented by the basis vectors included in the partial solution. Note that some of the quadrupole and octupole coefficients of the O_8 model are not well resolved, but included out of necessity (see discussion of parameter resolution in Connerney et al. (1991)).

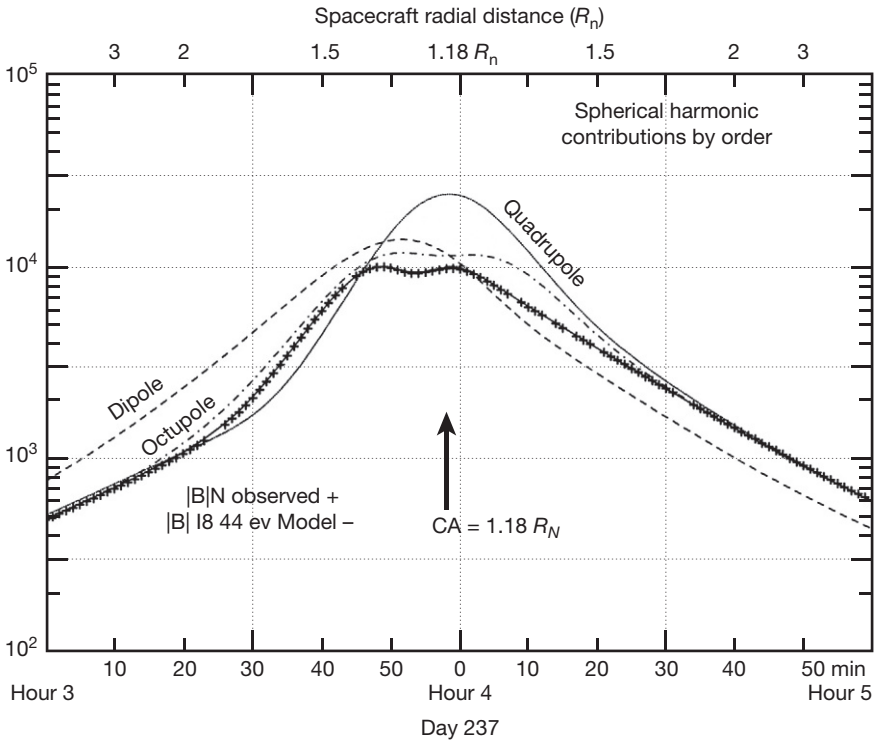


Figure 24 Magnitude of the observed magnetic field (crosses) as a function of time (lower axis) and spacecraft radial distance (upper axis) for the Voyager 2 encounter with Neptune, compared with a model fit (O_8 model) and the relative magnitudes of the dipole (dashed), quadrupole (solid), and octupole (dot-dashed) terms in the spherical harmonic expansion.

Table 7 Neptune magnetic field models

		Connerney et al. (1991)	*Holme and Bloxham (1996)*
Coefficients		*Neptune O_8*	*Neptune*
1	g_1^0	9732.0	10336.0
2	g_1^1	3220.0	3359.0
3	h_1^1	−9889.0	−9772.0
4	g_2^0	7448.0	8556.0
5	g_2^1	664.0	−406.0
6	g_2^2	4499.0	4644.0
7	h_2^1	11230.0	11139.0
8	h_2^2	−70.0	−743.0
9	g_3^0	−6592.0*	−5749.0
10	g_3^1	4098.0*	11632.0
11	g_3^2	−3581.0*	−1889.0
12	g_3^3	484.0	−2920.0
13	h_3^1	−3669.0	−3905.0
14	h_3^2	1791.0*	903.0
15	h_3^3	−770.0	−245.0

$1R_n = 24\,765$ km.
*Spherical harmonic coefficients in nT; coefficients that are *unresolved* are identified.

As anticipated, the relative magnitude of the quadrupole is quite large; the octupole is large as well, but can only be loosely constrained by the Voyager observations. The addition of a smoothness constraint (Holme and Bloxham, 1996) allows one to more confidently estimate the octupole coefficients, again resulting in a model that agrees well with the parameters in Connerney et al.'s (1991) model that are well resolved.

The magnetic field magnitude on the surface of the planet ranges from ~0.9 G at southern midlatitudes to ≤0.1 G near the equator and at northern midlatitudes (**Figure 25**). **Figure 25** also shows estimated auroral zones (stippled regions) that differ greatly in size as a result of the hemispherical asymmetry in field magnitudes. In this representation, it must be recognized that only the low-degree and low-order part of the field is portrayed on the surface of the planet, and the likely presence of higher-degree and higher-order terms may be expected to significantly alter the surface field. It is clear that the magnetic fields of Uranus and Neptune are quite similar in the range of surface field magnitudes, the tilt of the dipole, and the appearance of one strong pole at southern midlatitudes. Both Uranus and Neptune have extended weak-field regions, multiple-dip equators, and multipolar structure in evidence at the surface of the planet.

10.06.6 Satellites and Small Bodies

10.06.6.1 Moon

The Moon, with a radius of 1738 km and a mass of little more than 1% that of the Earth ($m/M_e = 0.0123$), is the largest satellite among the terrestrial planets. It is not so much smaller than Mercury (2439 km) in size, with a little more than one-third the volume of Mercury, but only about one-fifth of the mass. It is a differentiated object similar in composition to the Earth's mantle.

The Moon has been a frequent target of space exploration spanning nearly half a century, beginning with a series of robotic probes in the 1960s (Ranger series of the US space program and Luna and Zond series of the USSR program) and culminating in the first missions of human exploration of the Moon, the US Apollo program. For a brief moment in history, beginning with Apollo 11 on 20 July 1969, and ending with Apollo 17 on 14 December 1972, humans walked (and drove) on the surface of another world, deploying instruments and collecting soil and rock samples for return to Earth. Much of what we know of lunar magnetism evolved from study of the lunar samples returned from the Moon, combined with *in situ* observations collected both on the surface and in orbit about the Moon. Robotic scientific exploration of the Moon was rekindled decades after the Apollo program ended, beginning with the orbiting Clementine spacecraft in February and March of 1994. The Lunar Prospector (LP) spacecraft was in polar mapping orbit from 13 January 1998 to 31 July 1999, mapping crustal magnetic fields via both direct (magnetometry) and indirect (electron reflectometry) methods. Initially, LP operated in polar orbit at a nominal 100 km mapping altitude, followed by an extended mission (January–July of 1999) at reduced altitudes of between 15 and 45 km. ESA's SMART-1 spacecraft was in orbit about the Moon from 15 November 2004 to impact on 3 September 2006. In October 2007, Japan's SELENE spacecraft was inserted into an ~100 km altitude, polar mapping orbit to map crustal magnetic anomalies, characterize the lunar plasma environment, and probe its electrical conductivity structure (Tsunakawa et al., 2010). SELENE was a product of the ISAS and the National Space Development Agency (NASDA), both now part of the Japan Aerospace Exploration Agency (JAXA). After 20 months in orbit, SELENE impacted the lunar surface near the crater Gill, having completed its mapping mission.

The first complete lunar magnetic field map (**Figure 26**) to utilize the electron reflectometry method was compiled using 1.5×10^5 estimates of the surface field magnitude obtained over LP's 18-month mission (Mitchell et al., 2008). This map extends the electron reflection technique, correcting for charging of the lunar surface and spacecraft (Halekas et al., 2002), significantly improving the dynamic range of field magnitudes mapped relative to previous maps (Halekas et al., 2001). The surface field magnitude is estimated with a resolution of about 0.2 nT in magnitude. This map is still undersampled, in the sense that substantially more observations would be required to obtain complete spatial coverage at the resolution implied by the electron gyroradius (about 5 km at 300 eV). However, it provides a good sense of the global distribution of magnetic sources in the lunar crust, and it demonstrates that the characteristic scale length of coherent magnetization in the lunar crust is quite small. Where the lunar crust is magnetized, it is weakly magnetized (on spatial scales of 5 km and greater), compared to either the Mars' crust or Earth's. Lunar crustal magnetization also appears to be less coherent at large spatial scales, compared to Earth and Mars. The late, large impact basins (e.g., Orientale and Imbrium) appear nonmagnetic, implying that the impact melt cooled in zero field (no lunar dynamo at the time of impact); and yet, the regions of strongest field magnitude appear *antipodal* to the largest impact basins (Lin et al., 1988).

So it seems that lunar crustal magnetization and large impacts are related; perhaps impacts magnetize the crust not at the impact site, but on the opposite side of the Moon. One theory for how this might occur (Hood and Huang, 1991;

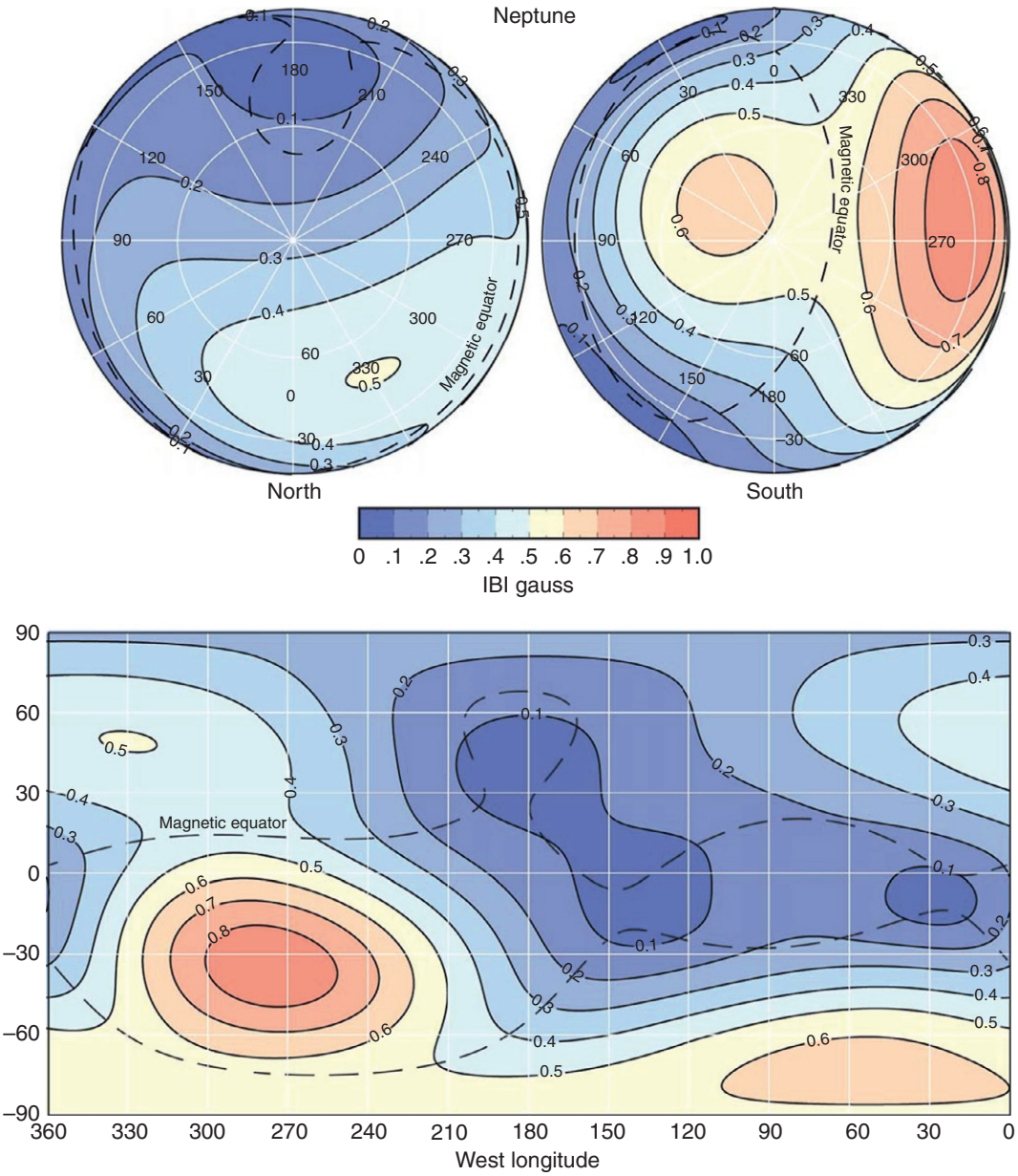


Figure 25 Contours of constant magnetic field magnitude (Gauss) on the dynamically flattened (1/47.6) surface of Neptune, computed using the O_8 spherical harmonic model (see text). The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) to $-z$ (right); below the color bar is a rectangular latitude–longitude (west longitude) projection. A trace (dashed) indicates the position of the magnetic equator.

Hood and Vickery, 1984) uses the impact-generated plasma cloud to intensify the weak ambient magnetic field at the antipode, if only briefly (~1 day), as it expands from the impact site to envelop the Moon. In this theory, the expanding plasma cloud, being a good electrical conductor, sweeps up the ambient field as it converges on the antipode, where the crust then acquires remanence via shock magnetization (Cisowski et al., 1983; Fuller et al., 1974). The efficacy of this process (the acquisition of remanence via shock in the presence of an applied magnetic field) has been demonstrated in shock experiments performed on lunar mare basalts (Gattacceca et al., 2010). Shock remanent magnetism (SRM) may be acquired by typical lunar rocks (dominated by multidomain iron–nickel mineralogy with low nickel content) subjected to low-pressure shocks (<10 GPa) attributed to lunar impacts. A detailed study of the magnetic field associated with multiring impact basins, estimated using the electron reflectometry technique, finds an extensive magnetic low of ~1.5–2 basin radii associated with each basin, attributed to shock demagnetization of the impact area (Halekas et al., 2003). A subset of these impact basins formed in the early Nectarian era (~3.9 Ga) also evidence a more compact, central magnetic high that is attributed to thermal remanence (TRM) in an ambient field, that is, evidence for a lunar dynamo operating at that time. The lack of such features in younger (Imbrian) impact basins (~3.2–3.85 Ga) and older (pre-Nectarian) impact basins (>3.92 Ga)

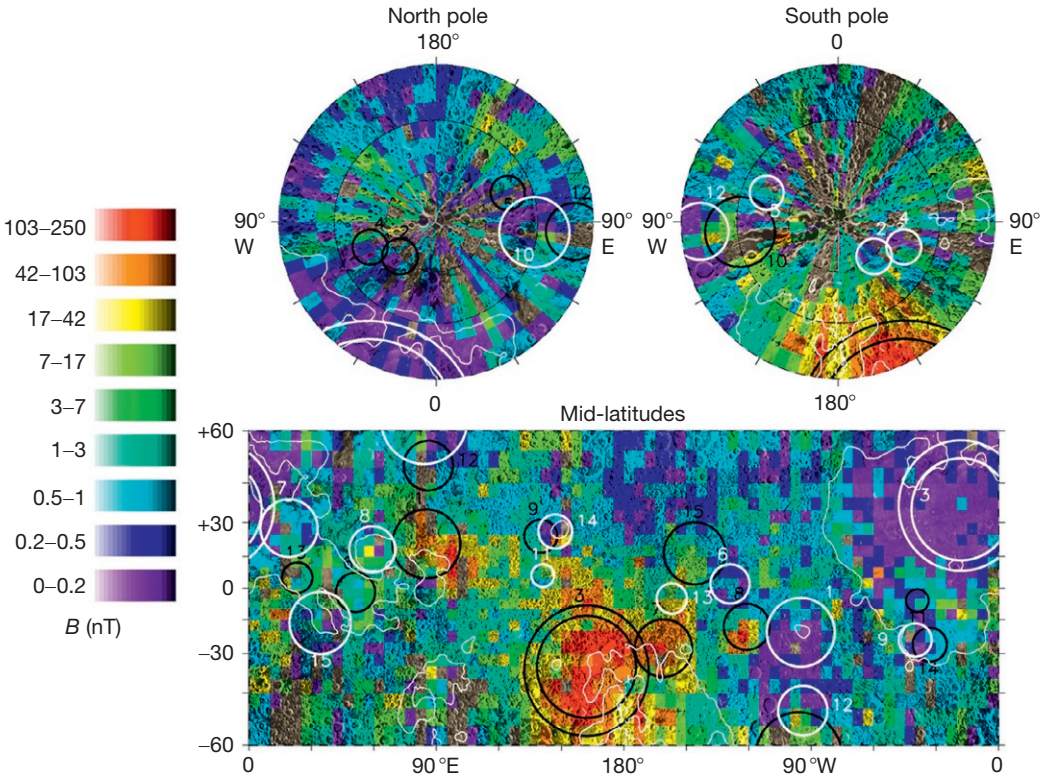


Figure 26 Magnetic field magnitude on the lunar surface as measured by the electron reflection technique. Field magnitudes from 0.2 to 250 nT are represented by colors according to the color bar to the left. The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) to $-z$ (right); below is a rectangular latitude–longitude projection. The white circles denote major impact basins, and the black circles are drawn at the antipodal site. Reproduced from Mitchell DL, Halekas JS, Lin RP, et al. (2008) Global mapping of lunar crustal magnetic fields by Lunar Prospector. *Icarus* 194: 401–409.

suggests that the lunar dynamo was active only briefly, for perhaps a few hundred million years during the period of time spanned by the (large) impact record. A detailed study of LP magnetometer observations acquired over these same basins at low altitude (~25–40 km) confirms the presence of central features in these Nectarian-aged basins (Hood, 2011). It is worthwhile to note that these features are very weak – just a few tenths of a nanotesla over most of the lunar surface – at satellite altitude. The most intense magnetic anomalies measured directly by the LP magnetometer range from a few to ~25 nT in magnitude (Hood et al., 2001; Richmond and Hood, 2008), sensed at ~15–20 km above the lunar surface. The most extensive magnetic anomalies are mapped antipodal to the Imbrium, Crisium, and Serenitatis basins and over Rimae Sirsalis and the Reiner Gamma formation. The ER observations sense (surface) field magnitudes that are about an order of magnitude greater, but even these estimates may be dramatically biased, that is, underestimated (Halekas et al., 2010), particularly if the anomalies sensed are of smaller spatial scale (<10 km) than sensed by the ~200 eV electrons utilized by the ER technique.

The loose association of lunar magnetic anomalies with impact site antipodes (Lin et al., 1988) has been the most perplexing aspect of lunar magnetization, in part due to the somewhat exotic nature of the proposed mechanism by which an impact on one side of the Moon magnetizes the crust on the opposite side (Hood and Huang, 1991; Hood and Vickery, 1984). Wieczorek et al. (2012) had proposed instead that the global distribution of lunar crustal magnetism is better understood as a consequence of the giant impact event that formed the largest and oldest impact crater on the Moon: that of the South Pole–Aitken basin. They modeled the oblique impact of an ~200 km diameter object and found a distribution of impactor mass emplaced on the lunar surface that resembles the distribution of the most intense lunar magnetic anomalies. They proposed that the magnetized material in the lunar crust was delivered by this differentiated impactor and acquired remanence by cooling in the presence of the Moon's early dynamo field. This proposal neatly circumvents the problem posed by the paltry thermoremanent susceptibility of typical lunar materials, substituting material from the impacting projectile with orders of magnitude greater thermoremanent susceptibility (Wieczorek et al., 2012). It is also consistent with the presence of widely scattered and isolated anomalies of small spatial scale, downrange of the South Pole–Aitken basin and elsewhere.

Complete global maps of the lunar magnetic field were also compiled using direct measurements obtained by the FGM (Purucker, 2008; Purucker and Nicholas, 2010; Richmond and Hood, 2008) at a nominal altitude of ~30 km provided by LP's extended mission. These maps (**Figure 27**) agree well with each other and with the map of field magnitudes derived from ER measurements if one allows for the differences in mapping altitudes. All evidence, to some degree, the broad

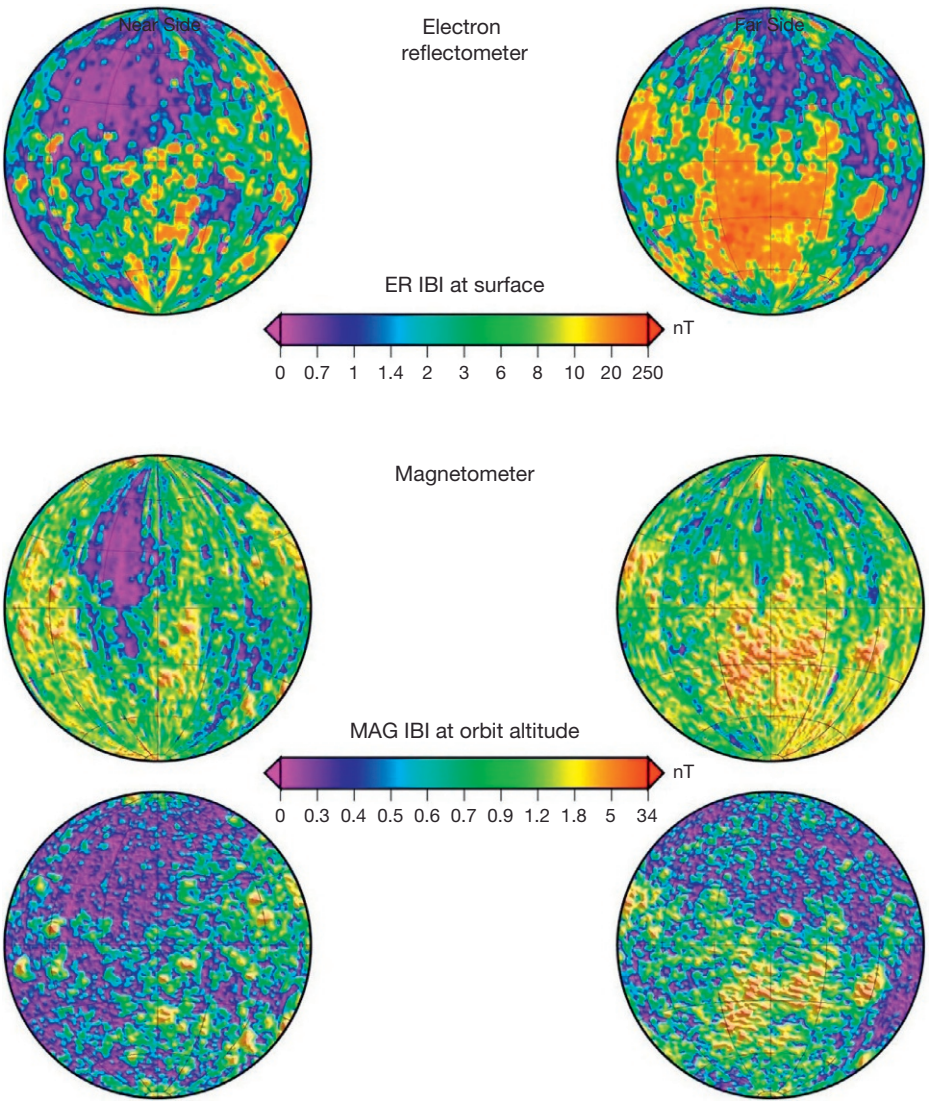


Figure 27 Lambert equal-area projections (nearside: left, farside: right) of the magnitude of the magnetic field above the lunar surface measured by the electron reflectometer (ER) and magnetometer instruments on Lunar Prospector (LP). The ER map (top) shows estimated field magnitudes at the lunar surface, remotely sensed via the electron reflection technique (smoothed over $3^\circ \times 3^\circ$ in latitude and longitude) (Mitchell et al., 2008). The LP magnetometer observations are represented in two forms. The middle set is color-coded to the magnitude of the magnetic field observed at ~30 km altitude after applying a correction for external fields (Richmond and Hood, 2008). The bottom set represents the field magnitude (also at 30 km altitude) derived from a 170° spherical harmonic model fitted to LP magnetometer data (Purucker and Nicholas, 2010). All measurements were acquired by LP at low altitude and primarily on the Moon's nightside, including passages through the Earth's magnetotail, when magnetically quiet conditions were experienced.

weak field associated with Imbrium on the nearside and the relatively strong crustal fields northwest of the South Pole–Aitken basin on the farside. More subtle features associated with other impact basins can also be found, particularly if one uses the ER map (top) as a guide in interpretation of the others. Tsunakawa et al. had applied a regional mapping method to map lunar surface magnetic fields (Tsunakawa et al., 2014), using both LP and SELENE (Kaguya) observations obtained in low lunar orbit (<50 km altitude above the South Pole–Aitken basin). They obtained a rather remarkable correspondence between fields mapped to the lunar surface using the two (independent) data sets, for both weak and strong field regions alike. Weak fields are characterized by surface field magnitudes of a few nanoteslas, whereas the strong field region to the northwest of the South Pole–Aitken basin is characterized by small-scale features with surface fields of ~200 nT. Both regions evidence elongated features, most apparent in the vector components of the mapped field, suggesting lineated source material.

The Moon's crustal magnetism can be understood as a result of impact processes resulting in both SRM and TRM, with rather weak magnetization intensities of a few tenths to a few amperes per meter (Hood, 2011; Purucker and Nicholas, 2010) and with spatial scales at or below the mapping resolution currently available (~5 km). Surface field magnitudes of a few hundred nanoteslas have been detected by the ER method, but Halekas

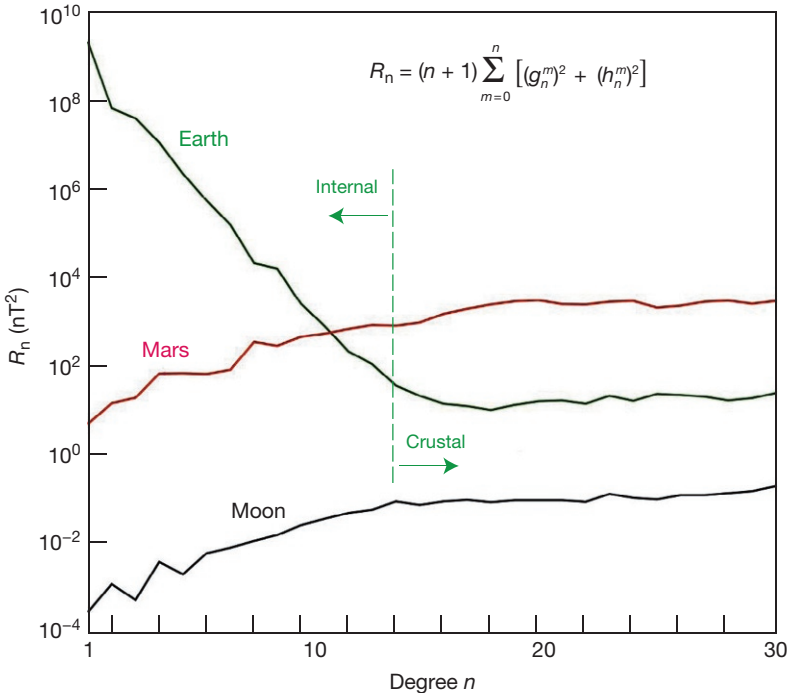


Figure 28 Relative harmonic content (Lowes' spectrum) of the magnetic fields of Earth, Moon, and Mars. The Earth's (internal) field at low degree is produced by the dynamo, but above $n=14$, crustal magnetic fields dominate the spectrum. At all spatial scales, the lunar crust is very weakly magnetized, compared to Earth and (especially) Mars.

et al. (2010) suggested that fields of order 1000 nT might be measured at the surface (with better resolution). One measure of the relative intensity of crustal magnetization is provided by Lowes' spectrum (**Figure 28**), R_n, which is equal to the mean-squared magnetic field amplitude attributed to harmonics of degree n. At all spatial scales measured, and modeled with the aid of spherical harmonics, the lunar crustal magnetic field is about an order of magnitude weaker than that of Earth ($n>14$) and two orders of magnitude weaker than that of Mars. This comparison likely underestimates the great differences in crustal magnetization among these bodies, however. The crustal fields of the Moon and Mars are due to remanent magnetism, whereas the Earth's crustal field is at least partly, if not mostly, induced. The lunar magnetic anomalies appear to be due to discrete magnetic sources scattered in an otherwise nonmagnetic (or very weakly magnetized) crust, whereas the Mars crust must be broadly magnetized throughout but for regions characterized by large impacts or volcanic emplacements.

10.06.6.2 Ganymede

Ganymede, with a radius of 2634 km, is the largest satellite in the solar system. It is slightly larger than the smallest planet in our solar system, Mercury ($R_m=2439$ km), but only about half as massive ($M_g=0.025M_e$). Ganymede may have an Fe or FeS core extending to as much as half the satellite's radius, overlain by a silicate mantle and topped with ~800 km layer of ice (Anderson et al., 1996; Schubert et al., 1996).

The Galileo spacecraft was targeted for several close flybys of Ganymede, the first of which occurred in June of 1996. The first two flyby encounters, with minimum altitudes of 838 and 264 km, demonstrated conclusively the presence of an intrinsic dipolar magnetic field and associated magnetosphere (Gurnett et al., 1996; Kivelson et al., 1996, 1997). Initial analysis of the magnetometer observations revealed a dipole field with equatorial surface field of 750 nT, nearly (within 10°) aligned with the rotation axis. A field of this magnitude is sufficient to produce a Ganymede 'minimagnetosphere' with all of the characteristics normally associated with a trapped particle population. Ganymede's immediate environment (**Figure 29**) includes field lines with both ends anchored within Ganymede, hosting a local trapped radiation environment, 'open' field lines that extend to the Jovian ionosphere, and Jovian field lines. A more comprehensive analysis, performed after the discovery of induced magnetic fields associated with Europa and Callisto (Khurana et al., 1998; Kivelson et al., 1999; Neubauer, 1998; Zimmer et al., 2000), allowed for the presence of an induced field in addition to the main field (Kivelson et al., 2002). This analysis demonstrated that Ganymede has both an inductive response to the time-varying magnetic field, due to currents induced in a hidden global ocean, and an intrinsic field likely due to a dynamo in the deep interior. The intrinsic field is a permanent dipole with a surface equatorial field of 719 nT, tilted by 176° with respect to the rotation axis and rotated 24° toward the trailing hemisphere from the Jupiter-facing meridian plane. The permanent field is perhaps the best evidence that Ganymede has a substantial Fe or FeS core in the interior (Schubert et al., 1996); dynamo action implies that it is fluid, and convecting, with sufficient energy input to sustain the dynamo. Zhan and Schubert (2012) constructed Ganymede dynamo models in which inner core convection is driven by Fe snow and FeS flotation, in addition to

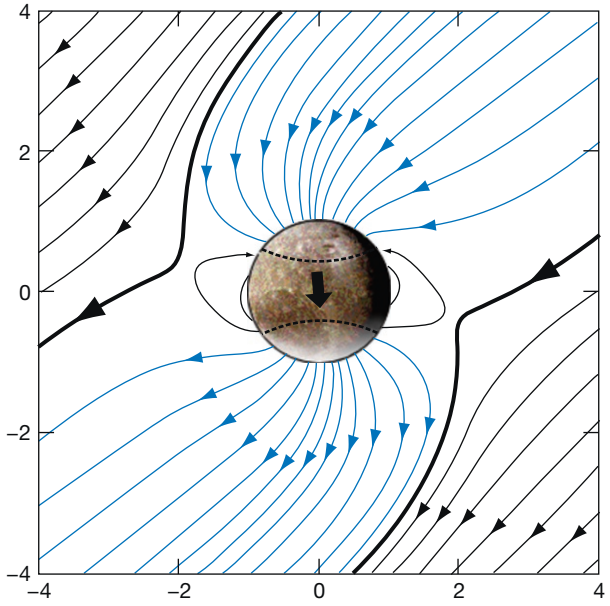


Figure 29 Magnetic field of Ganymede and its immediate environment in meridian plane projection. Ganymede's internal magnetic field moment is indicated with a filled arrow. Low-latitude field lines close within the satellite's interior; high-latitude field lines (blue) form a flux tube with one end rooted in Jupiter's ionosphere; Jovian field lines (black) are separated from those linking Ganymede by a separatrix (heavy black). Adapted from Kivelson MG, Khurana K, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.

solid inner core growth. They found that the Fe-snow mechanism is an unlikely candidate for powering Ganymede's dynamo, as it results in a magnetic field that is both weaker and more complex (multipolar) than is observed. Thus, core convection is mainly driven by FeS flotation or inner core growth, both of which concentrate the buoyancy flux near the inner core boundary, where it is most effective (Zhan and Schubert, 2012). However, it is possible (but not likely) that Ganymede's dipolar magnetic field is due to remanent magnetism (Crary and Bagenal, 1998).

10.06.6.3 Io

Io, with a radius of 1821 km, is the most geologically active satellite in the solar system. It is substantially smaller than the smallest magnetic planet in our solar system, Mercury ($R_m = 2439$ km), and less than half as massive as the largest satellite, Ganymede ($M_{io} = 0.39 M_g$). It is quite similar in size and mass to the Moon. Yet it is the most volcanic body in the solar system, evidencing extensive and ongoing volcanic activity owing to tidal heating. Io is a differentiated body with an Fe or FeS core extending to as much as one-third or one-half the satellite's radius, overlain by a silicate mantle and topped with a young, pockmarked crust a few tens of km thick (Anderson et al., 2001).

The Galileo spacecraft executed a number of close flybys of Io, probing fields in Io's immediate vicinity when it was both above and below the magnetodisk current sheet, where it experiences the maximum inducing field. Io, like the icy satellites Europa, Ganymede, and Callisto, evidences an inductive response to the time-varying magnetic field, due to currents flowing in an electrically conductive interior (Khurana et al., 2011). But in contrast to the icy satellites, Io's currents flow in a global subsurface magma layer just beneath the crust, rock heated to ~1200 °C by tidal heating caused by its orbital resonance with Europa and Ganymede (Peale et al., 1979). Khurana et al. (2011) considered the possibility that Io (like Ganymede) sustained a weak dynamo in addition to an inductive response, finding an upper limit of ~100 nT for a surface (equatorial) field, about 20% of the magnitude of the maximum inductive response. They concluded that if Io has a permanent dipole field, it is extremely weak. Why would Io, with a fluid, electrically conductive Fe or FeS core, and prodigious heat flow, not sustain an active dynamo? A likely explanation has to do with where the heat originates. Tidal heating deposits heat just beneath the crust, inhibiting transport of heat from the deep interior (Schubert and Soderlund, 2011; Weinbruch and Spohn, 1995). Convection in the core (and an active dynamo) cannot be maintained absent a temperature differential across the core–mantle boundary.

10.06.7 Discussion

A convenient measure of the complexity of a planetary magnetic field, often used in studies of the magnetic field of the earth and planets, is the 'harmonic spectrum,' sometimes referred to as a 'Lowes' spectrum,' defined as follows (Langel and Estes, 1982; Lowes, 1974):

$$R_n = (n+1)\sum_{m=0}^{n}\left\{\left(g_n^m\right)^2 + \left(h_n^m\right)^2\right\}$$

This quantity is equal to the mean-squared magnetic field intensity over the planet's surface produced by harmonics of degree n. Scaled to the core–mantle boundary with the factor $(a/r_c)^{2n+4}$, this quantity represents the mean-squared magnetic field intensity at the dynamo surface, one of several measures of interest in dynamo theory (e.g., Rädler and Ness, 1990; Schulz and Paulikas, 1990). The Earth's field is well known to high degree and order ($N_{max} = 23$). Scaled to the core–mantle boundary, the spectrum becomes almost flat for $n \leq 14$, suggesting a 'white' spectrum for the dynamo at the core–mantle boundary (e.g., Lowes, 1974). (The Earth's spectrum beyond $n \sim 14$ is attributed to crustal sources.) It is assumed that the core–mantle boundary, the location of which is very accurately known, represents the outer boundary of the geodynamo. Thus, it has often been assumed (prior to knowledge of Uranus' and Neptune's magnetic fields) that a white spectrum at the core–mantle boundary is a common feature of all planetary dynamos (Elphic and Russell, 1978), although in the Earth's case, the quadrupole is considerably less than expected. In **Figure 30**, R_n is calculated using the GSFC 12/83 model for the Earth (Langel and Estes, 1985), the Z_3 model for Saturn (Connerney et al., 1982b), and the Q_3 model for the magnetic field of Uranus (Connerney et al., 1987). For Jupiter, both the O_4 (Acuña and Ness, 1976) and the O_6 models (Connerney, 1992) are shown, illustrating a range of possible values for the quadrupole and octupole. For Uranus, no estimate of the magnitude of

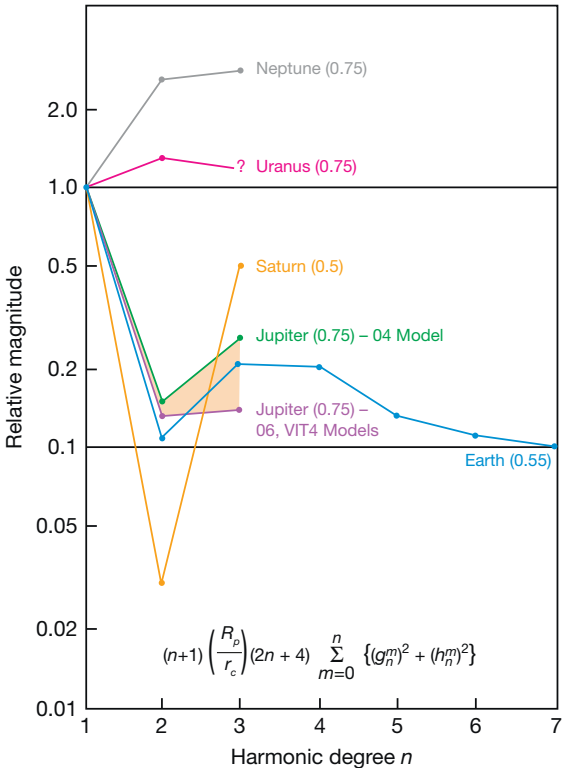


Figure 30 Relative harmonic content (Lowes' spectrum) of spherical harmonic models of Jupiter, Saturn, Uranus, and Neptune, compared with that of Earth, normalized to the assumed core radius (in parentheses) for each planet. Two Jovian magnetic field models are chosen to illustrate the range of possibilities for that planet.

higher-degree ($n>2$) moments can be made, from magnetic measurements alone (but see Holme and Bloxham, 1996 and Herbert, 2009), while for Neptune, terms of higher degree are appreciable, but quite uncertain.

For Jupiter and Saturn, we assume that the dynamo core radius is near the radius of the metallically conducting hydrogen core, $\sim$0.75 and 0.5 planet radius, respectively (e.g., Stevenson, 1983). The transition from molecular hydrogen to the metallic state is likely to be a gradual transition, and for Jupiter, especially, the molecular hydrogen envelope at depth is likely to be electrically conductive and capable of dynamo generation (Hide, 1967; Smoluchowski, 1975; Stanley and Glatzmaier, 2010). So Jupiter's dynamo may extend to $\sim 0.85R_j$ or so. For Uranus and Neptune, we take $r_c = 0.75R_p$ with the expectation that the dynamo for these planets operates in the ice mantle (Hubbard and MacFarlane, 1980; Hubbard and Marley, 1989; Podolak and Reynolds, 1981; Podolak et al., 1991), where ionic conduction prevails (Smoluchowski and Torbett, 1981). The pressure-induced ionization of H_2O provides sufficient electrical conductivity for dynamo generation (Mitchell and Nellis, 1982; Nellis et al., 1988) although it is orders of magnitude less conductive than an iron-rich terrestrial core or a metallically conducting hydrogen core.

In this presentation, the planets fall into two distinct groups, one in which the core surface field is dominated by the dipole (Earth, Jupiter, and Saturn) and one in which the core surface field is dominated by higher-degree contributions (Uranus and Neptune). However, in recognition of the peculiarity of Saturn's axisymmetric field, one might better think in terms of three classifications. In Stevenson's (1982) model, differential rotation of an outer conductive shell attenuates the nonaxisymmetric components of the dynamo within, thereby reducing the observed tilt by an order of magnitude or more from the 'nominal' value. This model has an appealing physical basis (material properties of the H–He fluid) and appears to satisfy other constraints, for example, helium depletion of the gaseous outer envelope and a relatively high planetary heat flux. The dipole-dominated group includes the planets with near alignment of dipole and rotation axes (≤ 10); the other includes the oblique rotators Uranus (59°) and Neptune (47°), the planets with a very large angular separation of the dipole and rotation axes. **Table 8** provides a comparison of the gross characteristics of the magnetic fields of the planets and satellites with active dynamos, using the eccentric dipole representation (Connerney, 1993). The eccentric dipole is computed from a spherical harmonic model and provides a useful characterization of the field at a distance, the dipole magnitude, tilt, and offset. Estimated minimum and maximum surface fields are computed using spherical harmonic representations, taking into account dynamical flattening, which is most significant for the gas giants.

10.06.8 Summary

The magnetic planets (and satellites) of the solar system provide a rich diversity of objects of relevance to dynamo theory. Mercury, Mars, Earth, and Ganymede have substantial iron cores with varying amounts of S or some other alloying elements; each has, or once had, an active dynamo. In this group, Mars and Earth appear similar, although the Mars dynamo is no longer active, and we have only the crustal remanence that remains today as evidence of its dynamo. Mercury and Ganymede are similar in size as well as magnetic properties; both have diminutive dynamos often considered so weak that alternatives to an active dynamo are proposed (Aharonson et al., 2004; Crary and Bagenal, 1998; Giampieri and Balogh, 2002). Mercury's weak field has been attributed to a dynamo operating beneath a stably stratified and electrically conductive shell (Christensen, 2006) that serves to attenuate the field or, alternately, a thin shell dynamo operating above a solid and electrically conductive inner core (Stanley et al., 2005). Both Mercury and Ganymede have dipolar or nearly dipolar fields more or less aligned with the spin axis.

Unmagnetized Venus reminds us that in dynamo theory, the 'devil is in the details.'

In Jupiter and Saturn, we have two planets of nearly the same composition and comparable size but with magnetic fields that are worlds apart. Jupiter's dynamo produces a magnetic field resembling Earth's, while Saturn's is uniquely spin-axisymmetric. Stevenson's account of Saturn's internal structure and magnetic field is appealing, as it provides a vehicle for understanding the differences between these gas giant planets. In Uranus and Neptune, we have again two planets of similar composition and size, this time with similar magnetic fields (fortunately!). These icy twins form a category of 'oblique rotators' with large dipole tilts and eccentric magnetic fields.

Table 8 Summary of planetary and satellite magnetic fields

	Mercury	*Earth*	*Jupiter*	*Saturn*	*Uranus*	*Neptune*	*Ganymede*
M (Gauss)	0.002	0.306	4.30	0.215	0.228	0.142	0.007
θ_d (°)	0.0	11.2	9.4	0.0	58.6	46.9	0.0
ϕ_d (°)	n/a	70.8	200.1	n/a	53.6	72.0	n/a
$\|r\|$ (R_p)	0.17	0.076	0.119	0.038	0.352	0.485	–
$\|B_s\|$ (Gauss) Min–max	0.0025–0.007	0.24–0.66	4.0–13.0	0.18–0.84	∼0.1–1.0	∼0.1–0.9	0.007–0.015
Oblateness	n/a	1/298	1/15.4	1/10.6	1/41.6	1/47.6	n/a
Model	(1)	GSFC1283	O_6	Z_3	Q_3	O_8	(2)

Dipole parameters are obtained from spherical harmonic models using the eccentric dipole representation (Connerney, 1993) unless otherwise noted; this is a good approximation for distant fields. The offset of the eccentric dipole ($|r|$) is given in units of the planet radius. The minimum and maximum magnetic field magnitude on the planet's surface ($|B_s|$) is obtained using the spherical harmonic model and oblateness noted. (1) Mercury model from Alexeev et al. (2010) and (2) Ganymede model from Kivelson et al. (2012).

Dynamo generation within these planets takes place in the relatively poorly conducting 'ice' mantle that sits atop a stably stratified and electrically conductive 'superionic' water interior (Redmer et al., 2011). Dynamos operating in spherical shells can produce dipole fields with large dipole tilts (Aubert and Wicht, 2004) and the harmonic content characteristic of ice giant dynamos (Stanley and Bloxham, 2006), but none thus far mimic the hemispheric asymmetry observed of such planetary fields. Heuristically, it is tempting to associate dipole dominance with small dipole tilt and dipole deficiency with large dipole tilt; it appears that dynamos operating in less conductive fluids are in the latter group. If dynamos operate with roughly comparable magnetic Reynolds numbers and convective velocities, those operating in poorly conductive fluids will require relatively large spatial scale lengths (Connerney, 1993; Connerney et al., 1987). If so, an ice giant dynamo, characterized by a relatively small number of large convective cells, might be expected to have a large dipole tilt (e.g., Elsasser, 1946; Parker, 1969).

In little more than three decades, we have sampled the magnetic fields of every planet in the solar system, some several times over. This diverse sample set challenges our understanding of the generation of magnetic fields in planetary interiors. Dynamos are ubiquitous and evidently remarkably responsive to subtle differences in interior composition, heat flow, and convection and as such offer clues to a planet's interior structure and evolution. For the present, dynamo theory matures on the heels of such discoveries (Schubert and Soderlund, 2011; Stanley and Glatzmaier, 2010); for the near future, we may all look forward to the global mapping of Jupiter's magnetic field in the coming years and that of Saturn, Mercury, and Ganymede as well. Remote Uranus and Neptune may hold onto their secrets a bit longer.

References

Acuña MH, Behannon KW, and Connerney JEP (1983a) Jupiter's magnetic field and magnetosphere. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 1–50. New York: Cambridge University Press.

Acuña MH, Connerney JEP, and Ness NF (1983b) The Z_3 zonal harmonic model of Saturn's magnetic field: Analyses and implications. *Journal of Geophysical Research* 88: 8771–8778.

Acuña MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.

Acuña MH, Connerney JEP, Wasilewski P, et al. (1992) Mars observer magnetic fields investigation. *Journal of Geophysical Research* 97: 7799–7814.

Acuña MH, Connerney JEP, Wasilewski P, et al. (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars Global Surveyor Mission. *Science* 279: 1676–1680.

Acuña MH and Ness NF (1976) Results from the GSFC fluxgate magnetometer on Pioneer 11. In: Gehrels T (ed.) *Jupiter*, pp. 830–847. Tucson, AZ: University of Arizona Press.

Acuña MH and Ness NF (1980) The magnetic field of Saturn: Pioneer 11 observations. *Science* 207: 444–446.

Acuña MH, Ness NF, and Connerney JEP (1980) The magnetic field of Saturn: Further studies of the Pioneer 11 observations. *Journal of Geophysical Research* 85: 5675–5678.

Aharonson O, Zuber MT, and Solomon SC (2004) Crustal remanence in an internally magnetized non-uniform shell: A possible source for Mercury's magnetic field? *Earth and Planetary Science Letters* 218: 261–268.

Albee AL, Palluconi FD, and Arvidson RE (1998) Mars global surveyor mission: Overview and status. *Science* 279(5357): 1671–1672.

Alexeev II, Belenkaya ES, Slavin JA, et al. (2010) Mercury's magnetospheric magnetic field after the first two MESSENGER flybys. *Icarus* 209: 23–39.

Anderson BJ, Acuña MH, Korth H, et al. (2008) The structure of mercury's magnetic field from MESSENGER's first flyby. *Science* 321: 82. http://dx.doi.org/10.1126/science.1159081.

Anderson BJ, Acuña MH, Korth H, et al. (2010) The magnetic field of Mercury. *Space Science Reviews* 152: 307. http://dx.doi.org/10.1007/s11214-009-9544-3.

Anderson JD, Jacobson RA, Lau EL, Moore WB, and Schubert G (2001) Io's gravity field and interior structure. *Journal of Geophysical Research* 106(E12): 32963–32969.

Anderson BJ, Johnson CL, Korth H, et al. (2011) The global magnetic field of mercury from MESSENGER orbital observations. *Science* 333: 1859. http://dx.doi.org/10.1126/science.1211001.

Anderson BJ, Johnson CL, Korth H, et al. (2012) Low-degree structure in Mercury's planetary magnetic field. *Journal of Geophysical Research* 117: E00L12. http://dx.doi.org/10.1029/2012JE004159.

Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996) Gravitational constraints on the internal structure of Ganymede. *Nature* 384(6609): 541–543.

Anderson JD and Schubert G (2007) Saturn's gravitational field, internal rotation, and interior structure. *Science* 317(5843): 1384–1387.

Anderson KA, Lin RP, McGuire RE, and McCoy JE (1975) Measurement of lunar and planetary magnetic fields by reflection of low energy electrons. *Space Science Instrumentation* 1: 439–470.

Anselmi A and Scoon GEN (2001) BepiColombo, ESA's Mercury cornerstone mission. *Planetary and Space Science* 49(14–15): 1409–1420.

Arkani-Hamed J (2001a) A 50-degree spherical harmonic model of the magnetic field of Mars. *Journal of Geophysical Research* 106(E10): 23197–23208.

Arkani-Hamed J (2001b) Paleomagnetic pole positions and pole reversals on Mars. *Journal of Geophysical Research* 28(17): 3409–3412.

Arkani-Hamed J (2002) An improved 50-degree spherical harmonic model of the magnetic field of Mars derived from both high-altitude and low-altitude data. *Journal of Geophysical Research* 107(E5). http://dx.doi.org/10.1029/2001JE001496.

Arkani-Hamed J and Boutin D (2004) Paleomagnetic poles of Mars: Revisited. *Journal of Geophysical Research* 109, E03011. http://dx.doi.org/10.1029/2003JE002229.

Aubert J and Wicht J (2004) Axial vs. equatorial dipolar models with implications for planetary magnetic fields. *Earth and Planetary Science Letters* 221: 409–419.

Backus GE (1968) Kinematics of secular variation in a perfectly conducting core. *Philosophical Transactions of the Royal Society of London* 263: 239–266.

Backus GE (1986) Poloidal and toroidal fields in geomagnetic field modeling. *Reviews of Geophysics* 24: 75–109.

Balogh A, Dougherty MK, Forsythe RJ, et al. (1992) Magnetic field measurements in the vicinity of Jupiter during the Ulysses flyby. *Science* 257: 1515–1518.

Benton ER (1979) Magnetic probing of planetary interiors. *Physics of the Earth and Planetary Interiors* 20: 111–118.

Benton ER and Muth LA (1979) On the strength of electric currents and zonal magnetic fields at the top of the Earth's core: Methodology and preliminary estimates. *Physics of the Earth and Planetary Interiors* 20: 127–133.

Bergan S and Engle IM (1981) Mercury magnetosphere and magnetotail revisited. *Journal of Geophysical Research: Space Physics* 86(NA3): 1617–1620.

Berge GL and Gulkis S (1976) Earth-based radio observations of Jupiter: Millimeter to meter wavelengths. In: Gehrels T (ed.) *Jupiter*, pp. 621–692. Tucson, AZ: University of Arizona Press.

Bigg EK (1964) Influence of the satellite Io on Jupiter's decametric emission. *Nature* 203: 1008–1010.

Blakely RJ (1995) *Potential Theory in Gravity and Magnetic Applications*. New York: Cambridge University Press.

Bolton SJ and The Juno Science Team (2010) The Juno mission. In: *Proceedings of the International Astronomical Union Symposium, No. 269*, pp. 92–100.

Bonfond B, Grodent D, Gerard J-C, et al. (2009) The Io UV footprint: Location, inter-spot distances and tail vertical extent. *Journal of Geophysical Research* 114: A07224. http://dx.doi.org/10.1029/2009JA014312.

Boutin D and Arkani-Hamed J (2006) Pole wandering of Mars: Evidence from paleomagnetic poles. *Icarus* 181: 13–25.

Breuer D and Spohn T (2003) Early plate tectonics vs. single plate tectonics on Mars: Evidence from magnetic field history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072.

Burke BF and Franklin KL (1955) Observations of a variable radio source associated with the planet Jupiter. *Journal of Geophysical Research* 60: 213–217.

Burton ME, Dougherty MK, and Russell CT (2009) Model of Saturn's internal magnetic field based on Cassini observations. *Planetary and Space Science* 57: 1706–1713.

Burton ME, Dougherty MK, and Russell CT (2010) Saturn's internal planetary magnetic field. *Geophysical Research Letters* 37: L24105. http://dx.doi.org/10.1029/2010GL045148.

Cain JC, Ferguson BB, and Mozzoni D (2003) An $n = 90$ internal potential function of the Martian crustal magnetic field. *Journal of Geophysical Research* 108(E2): 5008. http://dx.doi.org/10.1029/2000JE001487.

Cao H, Russell CT, Christensen UR, Dougherty MK, and Burton ME (2011) Saturn's very axisymmetric magnetic field: No detectable secular variation or tilt. *Earth and Planetary Science Letters* 304: 22–28.

Cao H, Russell CT, Wicht J, Christensen UR, and Dougherty MK (2012) Saturn's high degree magnetic moments: Evidence for a unique planetary dynamo. *Icarus* 221: 388–394.

Carr TD, Desch MD, and Alexander JK (1983) Phenomenology of magnetospheric radio emissions. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 226–284. New York: Cambridge University Press.

Cecconi B and Zarka P (2005) Model of a variable radio period for Saturn. *Journal of Geophysical Research* 110: A12203. http://dx.doi.org/10.1029/2005JA011085.

Chapman S and Bartels J (1940) *Geomagnetism*. New York: Oxford University Press, pp. 639–668.

Chenette DL and Davis L Jr. (1982) An analysis of the structure of Saturn's magnetic field using charged particle absorption signatures. *Journal of Geophysical Research* 87: 5267–5274.

Chicarro A, Martin P, and Trautner R (2004) The Mars Express mission: An overview. In: Wilson A (ed.) *ESA Special Publication-1240: Mars Express: A European Mission to the Red Planet*, pp. 3–13. Noordwijk: ESTEC.

Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058. http://dx.doi.org/10.1038/nature05342.

Christensen UR and Wicht J (2008) Models of magnetic field generation in partly stable planetary cores: Applications to Mercury and Saturn. *Icarus* 196: 16–34.

Cisowski SM, Collinson DW, Runcorn SK, Stephenson A, and Fuller M (1983) A review of lunar paleointensity data and implications for the origin of lunar magnetism. *Journal of Geophysical Research Supplement* 88: A691–A704.

Clarke JT, Ajello J, Ballester GE, et al. (2002) Ultraviolet emissions from the magnetic footprints of Io, Ganymede, and Europa on Jupiter. *Nature* 415: 997–1000.

Clarke JT, Ballester G, Trauger J, et al. (1996) Far-ultraviolet imaging of Jupiter's aurora and the Io "footprint" with the Hubble Space Telescope Wide Field Planetary Camera 2. *Science* 274: 404–409.

Clarke JT, Grodent D, Cowley S, et al. (2005) Jupiter's aurorae. In: Bagenal F, Dowling TE, and McKinnon W (eds.) *Jupiter: The Planet, Satellites, and Magnetosphere*, pp. 639–671. Cambridge: Cambridge University Press.

Clarke JT, Prange R, and Ballester GE (1995) HST far-ultraviolet imaging of Jupiter during the impacts of Comet Shoemaker-Levy-9. *Science* 267(5205): 1302–1307.

Colin L (1983) Basic facts about Venus. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 10–26. Tucson, AZ: University of Arizona Press.

Connerney JEP (1981) The magnetic field of Jupiter: A generalized inverse approach. *Journal of Geophysical Research* 86: 7679–7693.

Connerney JEP (1986) Magnetic connection for Saturn's rings and atmosphere. *Journal of Geophysical Research* 13(8): 773–776.

Connerney JEP (1992) Doing more with Jupiter's magnetic field. In: Rucker HO, Bauer SJ, and Kaiser ML (eds.) *Planetary Radio Emissions III*, pp. 13–33. Vienna: Austrian Academy of Science.

Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98(E10): 18659–18679.

Connerney JEP (2013) Saturn's ring rain. *Nature* 496(7444): 178–179.

Connerney JEP and Acuña MH (1982) Jovimagnetic secular variation. *Nature* 297: 313–315.

Connerney JEP, et al. (2014) *The Juno magnetic field investigation*. *Space Science Reviews* (in preparation).

Connerney JEP, Acuña MH, and Ness NF (1981) Modeling the Jovian current sheet and inner magnetosphere. *Journal of Geophysical Research* 86: 8370–8384.

Connerney JEP, Acuña MH, and Ness NF (1982a) Voyager 1 assessment of Jupiter's planetary magnetic field. *Journal of Geophysical Research* 87: 3623–3627.

Connerney JEP, Acuña MH, and Ness NF (1984a) The Z_3 model of Saturn's magnetic field and the Pioneer 11 Vector Helium magnetometer observations. *Journal of Geophysical Research* 89: 7541–7544.

Connerney JEP, Acuña MH, and Ness NF (1987) The magnetic field of Uranus. *Journal of Geophysical Research* 92(A13): 15329–15336.

Connerney JEP, Acuña MH, and Ness NF (1991) The magnetic field of Neptune. *Journal of Geophysical Research* 96: 19023–19042.

Connerney JEP, Acuña MH, and Ness NF (1996) Octupole model of Jupiter's magnetic field from Ulysses observations. *Journal of Geophysical Research* 101: 27453–27458.

Connerney JEP, Acuña MH, Ness NF, et al. (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102(42): 14970–14975.

Connerney JEP, Acuña MH, Ness NF, and Satoh T (1998) New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *Journal of Geophysical Research: Space Physics* 103(A6): 11929–11939.

Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111(1–2): 1–32.

Connerney JEP, Acuña MH, Wasilewski PJ, et al. (2001) The global magnetic field of Mars and implications for crustal evolution. *Geophysical Research Letters* 28: 4015–4018.

Connerney JEP, Acuña MH, Wasilewski PJ, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.

Connerney JEP, Baron R, Satoh T, and Owen T (1993) Images of excited H_3^+ at the Foot of the Io Flux Tube in Jupiter's atmosphere. *Science* 262: 1035–1038.

Connerney JEP, Davis L Jr., and Chenette DL (1984b) Magnetic field models. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 354–377. Tucson, AZ: University of Arizona Press.

Connerney JEP and Desch MD (1992) Comment on "Evidence of Saturn's magnetic field anomaly from Saturnian kilometric radiation high frequency limit" by Galopeau et al. *Journal of Geophysical Research* 97: 8713–8717.

Connerney JEP and Ness NF (1988) Mercury's magnetic field and interior. In: Matthews MS, Chapman L, and Vilas F (eds.) *Mercury*, pp. 494–513. Tucson, AZ: University of Arizona Press.

Connerney JEP, Ness NF, and Acuña MH (1982b) Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* 298: 44–46.

Connerney JEP and Satoh T (2000) The H_3^+ ion: A remote diagnostic of the Jovian magnetosphere. *Philosophical Transactions of the Royal Society of London* 358: 2471–2483.

Cowley SWH, Wright DM, Bunce EJ, et al. (2006) Cassini observations of planetary-period magnetic field oscillations in Saturn's magnetosphere: Doppler shifts and phase motion. *Geophysical Research Letters* 33(7): Art. No. L07104.

Cowling TG (1937) The magnetic field of sunspots. *Monthly Notices of the Royal Astronomical Society* 94: 39–48.

Cowling TG (1957) The dynamo maintenance of steady magnetic fields. *Quarterly Journal of Mechanics and Applied Mathematics* 10: 129–136.

Crary FJ and Bagenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research* 103: 25757–25774.

Crisp JA, Adler M, Matijevic JR, et al. (2003) Mars exploration Rover mission. *Journal of Geophysical Research* 108(E12), Art. No. 8061.

Davis L Jr. and Smith EJ (1985) Comments on "The Z_3 model of Saturn's magnetic field and the Pioneer 11 Vector Helium magnetometer observations" *Journal of Geophysical Research* 90: 4461–4464, Connerney, Acuña, Ness (eds.).

Davis L Jr. and Smith EJ (1986) New models of Saturn's magnetic field using Pioneer 11 Vector Helium magnetometer data. *Journal of Geophysical Research* 91: 1373–1380.

Davis L Jr. and Smith EJ (1990) A model of Saturn's magnetic field based on all available data. *Journal of Geophysical Research* 95(A9): 15257–15261.

Desch MD (1982) Evidence for solar-wind control of Saturn radio-emission. *Journal of Geophysical Research: Space Physics* 87(1): 4545–4549.

Desch MD, Connerney JEP, and Kaiser ML (1986) The rotation period of Uranus. *Nature* 322(6074): 42–43.

Desch MD and Kaiser ML (1981) Voyager measurement of the rotation rate of Saturn's magnetic field. *Geophysical Research Letters* 8: 253–256.

Desch MD and Rucker HO (1983) The relationship between Saturn kilometric radiation and the solar wind. *Journal of Geophysical Research* 88: 8999–9006.

Dessler AJ (1983) Coordinate systems. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 498–504. New York, NY: Cambridge University Press.

Dougherty MK, Achilleos N, Andre N, et al. (2005) Cassini magnetometer observations during Saturn orbit insertion. *Science* 307: 1266–1270.

Dougherty MK, Balogh A, Southwood DJ, and Smith EJ (1996) Ulysses assessment of the Jovian planetary field. *Journal of Geophysical Research: Space Physics* 101(A11): 24929–24941.

Dougherty MK, Kellock S, Southwood DJ, et al. (2004) The Cassini magnetic field investigation. *Space Science Reviews* 114(1–4): 331–383.

Elphic RC and Russell CT (1978) On the apparent source depth of planetary magnetic fields. *Geophysical Research Letters* 5: 211–214.

Elsasser WM (1946) Induction effects in terrestrial magnetism. II. The secular variation. *Physical Review* 70: 202–212.

Fairen AG, Ruiz J, and Anguita F (2002) An origin for the linear magnetic anomalies on Mars through accretion of terrains: Implications for dynamo timing. *Icarus* 160: 220–223.

Fairfield DH (1980) A statistical determination of the shape and position of the geomagnetic neutral sheet. *Journal of Geophysical Research* 85: 775–780.

Frawley JJ and Taylor PT (2004) Paleo-pole positions from Martian magnetic anomaly data. *Icarus* 172(2): 316–327.

Fuller M, Rose F, and Wasilewski PJ (1974) Preliminary results of an experimental study of the magnetic effects of shocking lunar soil. *The Moon* 9: 57–61.

Galopeau PHM and Lecacheux A (2000) Variations in Saturn's radio rotation period measured at kilometric wavelengths. *Journal of Geophysical Research* 105: 13089–13101.

Galopeau P, Ortega-Molina A, and Zarka P (1991) Evidence of Saturn's magnetic field anomaly from Saturnian kilometric radiation high frequency limit. *Journal of Geophysical Research* 96: 14129–14140.

Galopeau P, Zarka P, and LeQueau D (1995) Source location of Saturn's kilometric radiation: The Kelvin-Helmholtz instability hypothesis. *Journal of Geophysical Research* 100: 26397–26410.

Gattacceca J, Boustie M, Hood L, et al. (2010) Can the lunar crust be magnetized by shock: Experimental groundtruth. *Earth and Planetary Science Letters* 299: 42–53.

Giampieri G and Balogh A (2002) Mercury's thermoelectric dynamo model. *Planetary and Space Science* 50: 757–762.

Giampieri G and Dougherty MK (2004) Rotation rate of Saturn's interior from magnetic field observations. *Geophysical Research Letters* 31(16): L16701.

Giampieri G, Dougherty MK, Russell CT, and Smith EJ (2005) Reply to comment by Kaiser, M. L. et al. on "Rotation rate of Saturn's interior from magnetic field observations". *Geophysical Research Letters* 32(2): L02202.

Giampieri G, Dougherty MK, Smith EJ, and Russell CT (2006) A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* 441(7089): 62–64.

Gold RE, Solomon SC, McNutt RL, et al. (2001) The MESSENGER mission to Mercury: Scientific payload. *Planetary and Space Science* 49: 1467–1479.

Goldreich P and Lynden-Beil D (1969) Io a Jovian unipolar inductor. *Astrophysical Journal* 156(1P1): 59–78.

Goldreich P and Peale SJ (1966) Spin-orbit coupling in the solar system. *Astronomical Journal* 71: 425–438.

Gosling JT, McComas DJ, Thomsen MF, Bame SJ, and Russell CT (1986) The warped neutral sheet and plasma sheet in the near-earth geomagnetic tail. *Journal of Geophysical Research* 91: 7093–7099.

Graf JE, Zurek RW, Eisen HJ, et al. (2005) The Mars Reconnaissance Orbiter Mission. *Acta Astronautica* 57(2–8): 566–578.

Grodent D, Bonfond B, Gerard J-C, et al. (2008) Auroral evidence of a localized magnetic anomaly in Jupiter's northern hemisphere. *Journal of Geophysical Research* 113: A09201. http://dx.doi.org/10.1029/2008JA013185.

Grodent D, Bonfond B, Radioti A, et al. (2009) Auroral footprint of Ganymede. *Journal of Geophysical Research* 114: A07212. http://dx.doi.org/10.1029/2009JA014289.

Grodent D, Gerard J-C, Gustin J, et al. (2006) Europa's FUV auroral tail on Jupiter. *Geophysical Research Letters* 33: L06201. http://dx.doi.org/10.1029/2005GL025487.

Gurnett DA, Kurth WS, Roux A, Bolton SJ, and Kennel CF (1996) Evidence for a magnetosphere at Ganymede from plasma wave observations by the Galileo spacecraft. *Nature* 384: 535–537.

Gurnett DA, Persoon AM, Kurth WS, et al. (2007) The variable rotation period of the inner region of Saturn's plasma disk. *Science* 316: 442–445.

Halekas JS, Lillis RJ, Lin RP, et al. (2010) How strong are lunar crustal magnetic fields at the surface?: Considerations from a reexamination of the electron reflectometry technique. *Journal of Geophysical Research* 115: E03006. http://dx.doi.org/10.1029/2009JE003516.

Halekas JS, Lin RP, and Mitchell DL (2003) Magnetic fields of lunar multi-ring impact basins. *Meteoritics and Planetary Science* 38(4): 565–578.

Halekas JS, Mitchel DL, Lin RP, et al. (2001) Mapping of crustal magnetic anomalies on the lunar near side by the Lunar Prospector electron reflectometer. *Journal of Geophysical Research* 106(E11): 27841–27852.

Halekas J, Mitchell DL, Lin RP, et al. (2002) Evidence for negative charging of the lunar surface in shadow. *Geophysical Research Letters* 29: 77–81.

Helled R, Schubert GS, and Anderson JD (2009) Jupiter and Saturn rotation periods. *Planetary and Space Science* 57: 1467–1473.

Herbert F (2009) Aurora and magnetic field of Uranus. *Journal of Geophysical Research* 114A: 11206. http://dx.doi.org/10.1029/2009JA014394.

Hess SLG, Bonfond B, Zarka P, and Grodent T (2011) Model of the Jovian magnetic field topology constrained by the Io auroral emissions. *Journal of Geophysical Research* 116: A05217. http://dx.doi.org/10.1029/2010JA016262.

Hide R (1967) On the dynamics of Jupiter's interior and the origin of his magnetic field. In: Hindmarsh W, et al. (eds.) *Magnetism and the Cosmos*, pp. 378–395. Edinburgh: Oliver and Boyd.

Holme R and Bloxham J (1996) The magnetic fields of Uranus and Neptune: Methods and models. *Journal of Geophysical Research* 101: 2177–2200.

Hood LL (2011) Central magnetic anomalies of Nectarian-aged lunar impact basins: Probable evidence for an early core dynamo. *Icarus* 211: 1109–1128.

Hood LL, Harrison KP, Langlais B, et al. (2010) Magnetic anomalies near Apollinaris Patera and the Medusae Fossae Formation in Lucas Planum, Mars. *Icarus* 208: 118–131.

Hood LL and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins – Two dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.

Hood LL, Richmond NC, Harrison KP, and Lillis RJ (2007) East–west trending magnetic anomalies in the Southern hemisphere of Mars: Modeling analysis and interpretation. *Icarus* 191: 113–131.

Hood LL and Vickery A (1984) Magnetic field amplification and generation in hypervelocity meteoroid impacts with application to lunar paleomagnetism. *Journal of Geophysical Research* 89: 211–223.

Hood LL and Zakharian A (2001) Mapping and modeling of magnetic anomalies in the northern polar region of Mars. *Journal of Geophysical Research: Planets* 106(E7): 14601–14619.

Hood LL, Zakharian A, Halekas JS, et al. (2001) Initial mapping and interpretation of Lunar crustal magnetic anomalies using Lunar Prospector magnetometer data. *Journal of Geophysical Research* 106(E11): 27825–27839.

Hubbard WB and MacFarlane J (1980) Structure and evolution of Uranus and Neptune. *Journal of Geophysical Research* 85: 225–234.

Hubbard WB and Marley MS (1989) Optimized Jupiter, Saturn, and Uranus interior models. *Icarus* 78: 102–118.

Jackson DJ and Beard DB (1977) Magnetic field of Mercury. *Journal of Geophysical Research: Space Physics* 82(19): 2828–2836.

Johnson CL, Purucker ME, Korth H, et al. (2012) MESSENGER observations of Mercury's magnetic field structure. *Journal of Geophysical Research* 117: E00L14. http://dx.doi.org/10.1029/2012JE004217.

Johnson TV, Yeates CM, and Young R (1992) Space Science Reviews volume on Galileo Mission overview. *Space Science Reviews* 60: 3–21.

Jurdy DM and Stefanick M (2004) Vertical extrapolation of Mars magnetic potentials. *Journal of Geophysical Research* 109(E10): E10005. http://dx.doi.org/10.1029/2004JE002277.

Kaiser ML and Desch MD (1982) Saturn kilometric radiation: Source locations. *Journal of Geophysical Research* 87: 4555–4559.

Kaiser ML, Desch MD, Acuña MH, Ness NF, and Connerney JEP (2005) Comment on "Rotation rate of Saturn's interior from magnetic field observations". *Geophysical Research Letters* 32(2): L02201. http://dx.doi.org/10.1029/2004GL021398.

Kaiser ML, Desch MD, Kurth WS, et al. (1984) Saturn as a radio source. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 378–415. Tucson, AZ: University of Arizona Press.

Kargel JS (2006) Enceladus: Cosmic gymnast, volatile miniworld. *Science* 311: 1389–1391.

Khurana KK, Jai X, Kivelson MG, et al. (2011) Evidence of a global magma ocean in Io's interior. *Science* 332: 1186–1189. http://dx.doi.org/10.1126/science.1201245.

Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.

Kirk RL and Stevenson DJ (1987) Hydromagnetic constraints on deep zonal flow in the giant planets. *Astrophysical Journal* 316: 836–846.

Kivelson MG (2006) Does Enceladus govern magnetospheric dynamics at Saturn? *Science* 311: 1392.

Kivelson MG, Khurana KK, Coroniti FV, et al. (1997) The magnetic field and magnetosphere of Ganymede. *Geophysical Research Letters* 24: 2155–2158.

Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157(2): 507–522.

Kivelson MG, Khurana KK, Russell C, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.

Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104: 4609–4626.

Kivelson MG, Khurana K, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.

Kobayashi D and Sprenke KF (2010) Lithospheric drift on early Mars: Evidence in the magnetic field. *Icarus* 210: 37–42.

Korth H, Anderson BJ, Acuña MH, et al. (2004) Determination of the properties of Mercury's magnetic field by the MESSENGER mission. *Planetary and Space Science* 52: 733–746.

Kurth WS, Lecacheux A, Averkamp TF, Groene JB, and Gurnett DA (2007) A Saturnian longitude system based on a variable kilometric radiation period. *Geophysical Research Letters* 34(2): L02201.

Lanczos C (1971) *Linear Differential Operations*. Princeton, NJ: D. Van Nostrand, 564 pp.

Langel RA (1987) The main field. In: Jacobs JA (ed.) *Geomagnetism*, vol. 1, pp. 249–513. New York: Academic.

Langel RA and Estes RH (1982) A geomagnetic field spectrum. *Geophysical Research Letters* 9: 250–253.

Langel RA and Estes RH (1985) The near-Earth magnetic field at 1980 determined from MAGSAT data. *Journal of Geophysical Research* 90: 2495–2509.

Langlais B and Purucker M (2007) A polar magnetic paleopole associated with Apollinaris Patera, Mars. *Planetary and Space Science* 55: 270–279.

Langlais B, Purucker ME, and Mandea M (2004) Crustal magnetic field of Mars. *Journal of Geophysical Research: Planets* 109(E2): E02008.

Lian Y and Showman AP (2010) Generation of equatorial jets by large scale latent heating on the giant planets. *Icarus* 207: 373–393.

Lillis RJ, Frey HV, Manga M, et al. (2008) An improved crustal magnetic field map of Mars from electron reflectometry: Highland volcano magmatic history and the end of the Martian dynamo. *Icarus* 194: 575–596.

Lillis RJ, Mitchell DL, Lin RP, Connerney JEP, and Acuña MH (2004) Mapping crustal magnetic fields at Mars using electron reflectometry. *Geophysical Research Letters* 31: L15702.

Lillis RJ, Robbins S, Manga M, Halekas JS, and Frey HV (2013) Time history of the Martian dynamo from crater magnetic field analysis. *Journal of Geophysical Research* 118: 1488–1511. http://dx.doi.org/10.1002/jgre.20105.

Lin RP, Anderson KA, and Hood LL (1988) Lunar surface magnetic field concentrations antipodal to young large impact basins. *Icarus* 74: 529–541.

Lin RP, Mitchel DL, Curtis DW, et al. (1998) Lunar surface magnetic fields and their interaction with the solar wind: Results from Lunar Prospector. *Science* 281: 1480–1484.

Lortz D (1972) A simple stationary dynamo model. *Zeitschrift für Naturforschung* 27: 1350–1354.

Lowes FJ (1974) Spatial power spectrum of the main geomagnetic field and extrapolation to the core. *Geophysical Journal of the Royal Astronomical Society* 36: 717–730.

Mayhew MA and Estes RH (1983) Equivalent source modeling of the core magnetic field using MAGSAT data. *Journal of Geomagnetism and Geoelectricity* 35(4): 119–130.

Melendo EG, Perez-Hoyos S, Sanchez-Lavega A, and Hueso R (2011) Saturn's zonal wind profile in 2004–2009 from Cassini ISS images and its long term variability. *Icarus* 215: 1467–1473.

Milbury C and Schubert G (2010) Search for the global signature of the Martian dynamo. *Journal of Geophysical Research* 115(E1): E10010. http://dx.doi.org/10.1029/2010JE003617.

Milbury C, Schubert G, Raymond CA, Smrekar SE, and Langlais B (2012) The history of Mars' dynamo as revealed by modeling magnetic anomalies near Tyrrhenus Mons and Syrtis Major. *Journal of Geophysical Research* 117(E1): E10007. http://dx.doi.org/10.1029/2012JE004099.

Mitchell DL, Halekas JS, Lin RP, et al. (2008) Global mapping of Lunar crustal magnetic fields by Lunar Prospector. *Icarus* 194: 401–409.

Mitchell DL, Lillis RJ, Lin RP, Connerney JEP, and Acuña MH (2007) A global map of Mars' crustal magnetic field based on electron reflectometry. *Journal of Geophysical Research: Planets* 112(E1): E01002.

Mitchell AC and Nellis WJ (1982) Equation of state and electrical conductivity of water and ammonia shocked to the 100 GPa (1 Mbar) pressure range. *Journal of Chemical Physics* 76: 6273–6281.

Morris D and Berge GL (1962) Measurement of the polarization and angular extent of the decimeter radiation from Jupiter. *Astrophysical Journal* 136: 276–282.

Nellis WJ, Hamilton DC, Holmes NC, et al. (1988) The nature of the interior of Uranus based on studies of planetary ices at high dynamic pressure. *Science* 240: 779–781.

Ness NF (1979a) The magnetic fields of Mercury, Mars and Moon. *Annual Review of Earth and Planetary Sciences* 7: 248–288.

Ness NF (1979b) Magnetic field of Mercury. *Physics of the Earth and Planetary Interiors* 20(2–4): 209–217.

Ness NF, Acuña MH, Behannon KW, et al. (1981) Magnetic field studies by Voyager 1: Preliminary results at Saturn. *Science* 212: 211–217.

Ness NF, Acuña MH, Behannon KW, et al. (1982) Magnetic field studies by Voyager 2: Preliminary results at Saturn. *Science* 215: 558–563.

Ness NF, Acuña MH, Behannon KW, et al. (1986) Magnetic fields at Uranus. *Science* 233: 85–89.

Ness NF, Acuña MH, Burgala LF, et al. (1989) Magnetic fields at Neptune. *Science* 246: 1473–1478.

Ness NF, Acuña MH, Lepping RP, et al. (1979a) Magnetic field studies at Jupiter by Voyager 1: Preliminary results. *Science* 204: 982–987.

Ness NF, Acuña MH, Lepping RP, et al. (1979b) Magnetic field studies at Jupiter by Voyager 2: Preliminary results. *Science* 206: 966–972.

Ness NF, Behannon KW, Lepping RP, and Whang YC (1975) Magnetic field of Mercury confirmed. *Nature* 255: 204–205.

Ness NF, Behannon KW, Lepping RP, and Whang YC (1976) Observations of Mercury's magnetic field. *Icarus* 28: 479–488.

Ness NF, Behannon KW, Lepping RP, Whang YC, and Schatten KH (1974a) Magnetic field observations near Mercury: Preliminary results from Mariner 10. *Science* 185: 151–160.

Ness NF, Behannon KW, Lepping RP, Whang YC, and Schatten KH (1974b) Magnetic field observations near Venus: Preliminary results from Mariner 10. *Science* 185: 151–160.

Neubauer FM (1998) The sub-Alfvenic interaction of the Galilean satellites with the Jovian magnetosphere. *Journal of Geophysical Research* 103(E9): 19843–19866.

Ng KH and Beard DB (1979) Possible displacement of Mercury dipole. *Journal of Geophysical Research: Space Physics* 84(NA5): 2115–2117.

Nimmo F (2000) Dike intrusion as a possible cause of linear Martian magnetic anomalies. *Geology* 28: 391–394.

Nimmo F (2002) Why does Venus lack a magnetic field? *Geology* 30(11): 987–990. http://dx.doi.org/10.1130/0091-7613(2002)030.

Nimmo F and Stevenson D (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11976.

Northrop TG and Connerney JEP (1987) A micrometeorite erosion model and the age of Saturn's rings. *Icarus* 70: 124–137.

O'Donoghue J, Stallard TS, Melin H, et al. (2013) The domination of Saturn's low-latitude ionosphere by ring "rain". *Nature* 496(7444): 193–195. http://dx.doi.org/10.1038/nature12049.

Parker EN (1969) The occasional reversal of the Earth's magnetic field. *Astrophysics Journal* 158: 815–827.

Parker RL (2003) Ideal bodies for Mars magnetics. *Journal of Geophysical Research* 108(E1): 5006. http://dx.doi.org/10.1029/2001JE001760.

Peale SJ, Cassen P, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894. http://dx.doi.org/10.1126/science.203.4383.892.

Pettengill GH and Dyce RB (1965) A radar determination of the rotation of the planet Mercury. *Nature* 206: 1240.

Phillips JL and Russell CT (1987) Upper limit on the intrinsic magnetic field of Venus. *Journal of Geophysical Research* 92: 2253–2263.

Podolak M, Hubbard WB, and Stevenson DJ (1991) Models of Uranus' interior and magnetic field. In: Bergstralh JT, Miner ED, and Matthews MS (eds.) *Uranus*, pp. 29–61. Tucson, AZ: University of Arizona Press.

Podolak M and Reynolds RT (1981) On the structure and composition of Uranus and Neptune. *Icarus* 46: 40–50.

Porco CC and Danielson GE (1982) The periodic variation of spokes in Saturn's rings. *Astrophysical Journal* 87: 826–833.

Prange R, Rego D, Southwood D, et al. (1996) Rapid energy dissipation and variability of the Io-Jupiter electrodynamic circuit. *Nature* 379: 232–235.

Purucker ME (2008) A global model of the internal magnetic field of the moon based on Lunar Prospector magnetometer observations. *Icarus* 197: 19–23.

Purucker ME and Nicholas JB (2010) Global spherical harmonic models of the internal magnetic field of the Moon based on sequential and coestimation approaches. *Journal of Geophysical Research* 115: E12007. http://dx.doi.org/10.1029/2010JE003650.

Purucker ME, Ravat D, Frey H, Sabaka T, and Acuña MH (2000) An altitude-normalized magnetic map of Mars and its interpretation. *Geophysical Research Letters* 27(16): 2449–2452.

Purucker ME, Sabaka TJ, Solomon SC, et al. (2009) Mercury's internal magnetic field: Constraints on large- and small-scale fields of crustal origin. *Earth and Planetary Science Letters* 285: 340–346.

Rädler K-H and Ness NF (1990) The symmetry properties of planetary magnetic fields. *Journal of Geophysical Research* 95(A3): 2311–2318.

Ravat D (2011) Interpretation of Mars southern highlands high amplitude magnetic field with total gradient and fractal source modeling: New insights into the magnetic mystery of Mars. *Icarus* 214: 400–412.

Read PL, Dowling TE, and Schubert G (2009) Saturn's rotation period from its atmospheric planetary-wave configuration. *Nature* 460: 608–610.

Redmer R, Mattsson TR, Nettelmann N, and French M (2011) The phase diagram of water and the magnetic fields of Uranus and Neptune. *Icarus* 211: 798–803.

Richmond NC and Hood LL (2008) A preliminary global map of the vector lunar crustal magnetic field based on Lunar Prospector magnetometer data. *Journal of Geophysical Research: Planets* 113: E02010. http://dx.doi.org/10.1029/2007JE002933.

Riddle AC and Warwick JW (1976) Redefinition of System III longitude. *Icarus* 27: 457–459.

Riedler W, Mohlmann D, Oraevsky VN, et al. (1989) Magnetic fields near Mars – 1st results. *Nature* 341(6243): 604–607.

Russell CT (1978a) The magnetic field of Mars: Mars 3 evidence re-examined. *Geophysical Research Letters* 5: 81–84.

Russell CT (1978b) The magnetic field of Mars: Mars 5 evidence re-examined. *Geophysical Research Letters* 5: 85–88.

Russell CT (1979) The interaction of the solar wind with Mars, Venus and Mercury. In: Kennel CF (ed.) *Solar System Plasma Physics*, pp. 208–252. Amsterdam: Elsevier.

Russell CT and Dougherty MK (2010) Magnetic fields of the outer planets. *Space Science Reviews* 152: 251–269. http://dx.doi.org/10.1007/s11214-009-9621-7.

Russell CT, Elphic RC, Luhmann JG, and Slavin JA (1980) On the search for an intrinsic magnetic field at Venus. *Proceedings of the Lunar and Planetary Science Conference* 11: 1897–1906.

Russell CT and Vaisberg O (1983) The interaction of the solar wind with Venus. In: Hunten M, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 873–940. Tucson, AZ: University of Arizona Press.

Sabaka TJ, Olsen N, and Langel RA (2002) A comprehensive model of the quiet-time, near-Earth magnetic field: Phase 3. *Geophysical Journal International* 151: 32–68.

Sandel BR, Shemansky DE, Broadfoot AL, et al. (1982) Extreme ultraviolet observations from the Voyager 2 encounter with Saturn. *Science* 215: 548–553.

Saunders RS, Arvidson RE, Badhwar GD, et al. (2004) 2001 Mars Odyssey mission summary. *Space Science Reviews* 110(1–2): 1–36.

Schubert G, Chan KH, Xinhao L, and Zhang K (2004) Planetary dynamos: Effects of electrical conducting flows overlying turbulent regions of magnetic field generation. *Icarus* 172: 305–315.

Schubert G, Russell CT, and Moore WB (2000) Geophysics – Timing of the Martian dynamo. *Nature* 408: 666–667.

Schubert G and Soderlund KM (2011) Planetary magnetic fields: Observations and models. *Physics of the Earth and Planetary Interiors* 187: 92–108.

Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep N (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.

Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.

Schubert G, Zhang K, Kivelson MG, and Anderson JD (1996) The magnetic field and internal structure of Ganymede. *Nature* 384: 544–545.

Schultz PH and Lutz AB (1988) Polar wandering on Mars. *Icarus* 73(1): 91–141.

Scuffham J, Giampieri G, and Balogh A (2006) BepiColombo measurements of Mercury's internal field. *Advances in Space Research* 38: 610–615.

Showman AP (2009) Windy clues to Saturn's spin. *Nature* 460: 582–583.

Simpson JA, Bastian TS, Chenette DL, Mckibben RB, and Pyle KR (1980) The trapped radiations of Saturn and their absorption by satellites and rings. *Journal of Geophysical Research* 85: 5731–5762.

Sleep NH (1990) Hotspots and mantle plumes – Some phenomenology. *Journal of Geophysical Research: Solid Earth and Planets* 95(B5): 6715–6736.

Sleep NH (1994) Martian plate tectonics. *Journal of Geophysical Research* 99: 5639–5655.

Smith EJ, Connor BV, and Foster GT Jr. (1975a) Measuring the magnetic fields of Jupiter and the outer solar system. *IEEE Transactions on Magnetics* MAG-11: 962–980.

Smith EJ, Davis L Jr., Coleman PJ, and Jones DE (1965a) Magnetic field measurements near Mars. *Science* 149: 1241–1242.

Smith EJ, Davis L Jr., Coleman PJ, and Sonnet CP (1963) Mariner II: Preliminary reports on measurements of Venus magnetic field. *Science* 139: 909–910.

Smith EJ, Davis L Jr., Coleman PJ, and Sonnet CP (1965b) Magnetic measurements near Venus. *Journal of Geophysical Research* 70: 1571–1586.

Smith EJ, Davis L Jr., and Jones DE (1976) Jupiter's magnetic field and magnetosphere. In: Gehrels T (ed.) *Jupiter*, pp. 788–829. Tucson, AZ: University of Arizona Press.

Smith EJ, Davis L Jr., Jones DE, et al. (1974) The magnetic field of Jupiter and its interaction with the solar wind. *Science* 183: 305–306.

Smith EJ, Davis L Jr., Jones DE, et al. (1975b) Jupiter's magnetic field, magnetosphere, and its interaction with the solar wind: Pioneer 11. *Science* 188: 451–455.

Smith EJ, Davis L Jr., Jones DE, et al. (1980a) Saturn's magnetic field and magnetosphere. *Science* 207: 407–410.

Smith EJ, Davis L Jr., Jones DE, et al. (1980b) Saturn's magnetosphere and its interaction with the solar wind. *Journal of Geophysical Research* 85: 5655–5674.

Smith BA, Soderblom L, Batson R, et al. (1982) A new look at the Saturn system: Voyager 2 images. *Science* 215: 505–537.

Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.

Smoluchowski R (1975) Jupiter's molecular hydrogen layer and the magnetic field. *Astrophysical Journal Letters* 200: 119–121.

Smoluchowski R and Torbett M (1981) Can magnetic fields be generated in the icy mantles of Uranus and Neptune? *Icarus* 48: 146–148.

Solomon SC, McNutt RL Jr., Gold RE, et al. (2001) The Messenger mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465.

Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90(2): 222–236.

Sprenke KF (2005) Martian magnetic paleopoles: A geostatistical approach. *Geophysical Research Letters* 32(9): L09201.

Sprenke KF and Baker LL (2000) Magnetization, paleomagnetic poles, and polar wander on Mars. *Icarus* 147: 26–34.

Stanley S (2010) A dynamo model for axisymmetrizing Saturn's magnetic field. *Geophysical Research Letters* 37: L05201. http://dx.doi.org/10.1029/2009GL041752.

Stanley S and Bloxham J (2006) Numerical dynamo models of Uranus' and Neptune magnetic fields. *Icarus* 184: 556–572.

Stanley S, Bloxham J, Hutchison WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.

Stanley S and Glatzmaier GA (2010) Dynamo models for planets other than Earth. *Space Science Reviews* 152: 617–649. http://dx.doi.org/10.1007/s11214-009-9573-y.

Stanley S and Mohammadi A (2008) Effects of an outer thin stable stratified layer on planetary dynamos. *Physics of the Earth and Planetary Interiors* 168: 179–190.

Sterenborg MG and Bloxham J (2010) Can Cassini magnetic field measurements be used to find the rotation period of Saturn's interior? *Geophysical Research Letters* 37: L11201. http://dx.doi.org/10.1029/2010GL043250.

Stevenson DJ (1980) Saturn's luminosity and magnetism. *Science* 208: 746–748.

Stevenson DJ (1982) Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophysical and Astrophysical Fluid Dynamics* 21: 113–127.

Stevenson DJ (1983) Planetary magnetic fields. *Reports on Progress in Physics* 46: 555–620.

Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.

Stevenson DJ (2006) A new spin on Saturn. *Nature* 441: 34–35.

Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.

Tsunakawa H, Shibuya H, Takahashi F, et al. (2010) Lunar magnetic field observation and initial global mapping of lunar magnetic anomalies by MAP-LMAG onboard SELENE. *Space Science Reviews* 154: 219–251. http://dx.doi.org/10.1007/s11214-010-9652-0.

Tsunakawa H, Takahashi F, Shimizu H, Shibuya H, and Matsushima M (2014) Regional mapping of the lunar magnetic anomalies at the surface: Method and its application to strong and weak magnetic anomaly regions. *Icarus* 228: 35–53. http://dx.doi.org/10.1016/j.icarus.2013.09.026.

Uno H, Johnson CL, Andersen BJ, Korth H, and Solomon S (2009) Modeling Mercury's internal magnetic field with smooth inversions. *Earth and Planetary Science Letters* 285: 328–339.

Van Allen JA, Thomsen MF, Randall BA, Rairden RL, and Grosskreutz CL (1980) Saturn's magnetosphere, rings, and inner satellites. *Science* 207: 415–421.

Warwick JW (1963) The position and sign of Jupiter's magnetic moment. *Astrophysical Journal* 137: 1317–1318.

Warwick JW, Evans DR, Peltzer GR, et al. (1989) Voyager planetary radio astronomy at Neptune. *Science* 246(4936): 1498–1501.

Weir ST, Mitchell AC, and Nellis WJ (1996) Metallization of fluid molecular Hydrogen at 140 GPa (1.4 Mbar). *Physical Review Letters* 76: 1860–1863.

Weinbruch U and Spohn T (1995) A self sustained magnetic field on Io?. *Planetary and Space Science* 43(9): 1045–1057.

Whang YC (1977) Magnetospheric magnetic field of Mercury. *Journal of Geophysical Research: Space Physics* 82(7): 1024–1030.

Wieczorek MA, Weiss BP, and Stewart ST (2012) An impactor origin for lunar magnetic anomalies. *Science* 335: 1212–1215. http://dx.doi.org/10.1126/science.1214773.

Yin A (2012) Structural analysis of the Valles Marineris fault zone: Possible evidence for large-scale strike-slip faulting on Mars. *Lithosphere* 4: 286–330. http://dx.doi.org/10.1130/L192.1.

Yu ZJ, Leinweber HK, and Russell CT (2009) Galileo constraints on the secular variation of the Jovian magnetic field. *Journal of Geophysical Research* 115: E03002. http://dx.doi.org/10.1029/2009JE003492.

Zhan X and Schubert G (2012) Powering Ganymede's dynamo. *Journal of Geophysical Research* 117: E08011. http://dx.doi.org/10.1029/2012JE004052.

Zhang TL, Baumjohann W, Delva M, et al. (2006) Magnetic field investigation of the Venus plasma environment: Expected new results from Venus Express. *Planetary and Space Science* 54: 1336–1343.

Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.

This page intentionally left blank

10.07 Planetary Dynamos

FH Busse, Physikalisches Institut der Universität Bayreuth, Bayreuth, Germany
RD Simitev, University of Glasgow, Glasgow, UK

© 2015 Elsevier B.V. All rights reserved.

10.07.1	**Historical Introduction**	239
10.07.2	**General Remarks on the Dynamo Theory of Planetary Magnetism**	240
10.07.3	**Mathematical Formulation of the Problem of Spherical Dynamos**	241
10.07.4	**Convection in Rotating Spherical Shells**	242
10.07.5	**Convection-Driven Dynamos**	245
10.07.6	**Applications to Planetary Dynamos**	248
10.07.6.1	General Considerations	248
10.07.6.2	Mercury	250
10.07.6.3	Venus	250
10.07.6.4	Mars	251
10.07.6.5	Moon	251
10.07.6.6	Jupiter	251
10.07.6.7	Saturn	251
10.07.6.8	Uranus and Neptune	251
10.07.6.9	Ganymede and Other Satellites	252
10.07.7	**Concluding Remarks**	252
References		252

10.07.1 Historical Introduction

While the Earth's magnetism has been studied for centuries starting with the first scientific monograph of Gilbert (1600), the question of the magnetism of other planets had received scant attention until recently because of the lack of relevant observations. Only in 1955, clear evidence for the existence of a planetary magnetic field other than the geomagnetic one was obtained through the observation of the Jovian decametric radio waves (Burke and Franklin, 1955). Since it had been more or less accepted until the end of the nineteenth century that geomagnetism arises from the remnant magnetization of the Earth, similar properties may have been assumed for the other terrestrial planets and the Moon. This view lost its appeal, however, when it became evident that the Curie temperature is exceeded in the Earth below a depth of about 30 km. Ferromagnetic materials in the Earth's crust could thus explain only magnetic fields with a relatively short wavelength.

The current period of intense research on the magnetism of planets other than that of the Earth started with the first detailed measurement of Jupiter's magnetic field by the Pioneer 10 space probe in 1973 and the discovery of Mercury's magnetism by Mariner 10 in 1974. In the early 1970s, also the development of the theory of magnetohydrodynamic (MHD) dynamos had started in which the reaction of the Lorentz force of the generated magnetic field is taken into account in physically realistic configurations (Childress and Soward, 1972). Until that time, dynamo theoreticians had focused their attention on the kinematic problem in which the possibility of growing magnetic fields driven by somewhat arbitrarily chosen velocity fields is considered. It must be remembered that only a few years earlier, it had been demonstrated in a mathematically convincing way by Backus (1958) and Herzenberg (1958) that the homogeneous dynamo process of the generation of magnetic fields in a singly connected domain of an electrically conducting fluid is indeed possible. Doubts about the feasibility of this process that had first been proposed by Larmor (1919) had persisted after Cowling (1934) had proved that purely axisymmetric fields could not be generated in this way.

The complexity of the MHD dynamo problem described by the nonlinearly coupled Navier–Stokes equations and the equation of magnetic induction had prevented progress in understanding planetary dynamos through analytic solutions. Only the advent of powerful enough computers in the 1990s has allowed to solve numerically the coupled three-dimensional partial differential equations through forward integration in time. Even today and for the foreseeable future, the limits of computer capacity will permit the exploration of only a small fraction of the parameter space of interest for the understanding of planetary dynamos (see also **Chapters 8.03** and **8.10**).

In view of the difficulties of a rigorous theory of planetary dynamos, many attempts have been made to obtain simple similarity relationships that would fit the observed planetary magnetic moments as function of certain properties of the planets. Some early proponents have gone as far as claiming the existence of a 'magnetic Bode's law' corresponding to a relationship between the magnetic moment and size or angular momentum of a planet in analogy to the Titius–Bode law for the radii of planetary orbits. Just as in the latter case, however, attempts to derive a magnetic Bode's law from basic physical principles have failed.

Other proposals have taken into account physical forces. Since a common ingredient of planetary dynamos is the existence of a fluid part of the core with a sufficiently high electrical conductivity, the latter parameter together with the core radius and the angular velocity of the planetary rotation usually enters

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00172-X

the similarity relationships such as those proposed by Hide (1974), Busse (1976), Jacobs (1979), and Dolginov (1977). Malkus (1968, 1994) had argued for the precession of the Earth as the cause of geomagnetism, and he and Vanyo (1984) had demonstrated through laboratory experiments that precession and tides may cause turbulent motions in fluid planetary cores. Dolginov (1977) proposed a scaling law for the precessional origin of all planetary magnetic fields. While a dynamo driven by turbulent flows caused by precession and tides cannot be easily excluded in the case of the Earth (Tilgner, 2007), it is much less likely in the case of other planets such as Uranus for which precessional torques are rather minute. Just as a common precessional origin of planetary magnetism is not regarded as feasible, so have all other proposed similarity relationships lost in appeal and are no longer seriously considered. We shall return, however, to scaling relationships based more directly on the basic MHD equations in Section 10.07.5.

10.07.2 General Remarks on the Dynamo Theory of Planetary Magnetism

Since the proposal of the geodynamo as the cause of the Earth's magnetism had been in doubt for a long time before 1958, numerous alternative proposals had been made in the literature. Among these, only the possibility that thermoelectric currents may generate a planetary magnetic field is still discussed in the case of Mercury (Giampieri and Balogh, 2002; Stevenson, 1987). For a discussion of the failings of the various proposals for nondynamo origins of planetary magnetic fields, we refer to the papers of Bullard (1949) and Stevenson (1983).

Dynamos generally convert mechanical energy into magnetic one. In contrast to the technical dynamo that is characterized by a multiply connected region of high electrical conductivity, that is, it depends on an appropriate wiring, planetary dynamos are referred to as homogeneous dynamos since they operate in a singly connected fluid domain of high electrical conductivity. Since flows in planetary cores with active dynamos are usually turbulent, the small-scale structure of the magnetic field is correspondingly chaotic. The large-scale structure, however, can be quite regular. One distinguishes 'steady' and oscillatory dynamos. The most famous example of the latter kind is the solar dynamo that exhibits a well-defined period of about 22 years. The geodynamo, on the other hand, is usually regarded as a 'steady' dynamo, even though it varies in its amplitude by a factor of two or more on the magnetic diffusion timescale and reverses its polarity on a longer timescale. For an alternative view, see Busse and Simitev (2008). A measure of the magnetic diffusion time is given by the decay time, $t_d = \sigma\mu r_0^2/\pi^2$, of the magnetic field in the absence of fluid motions. Here, σ and μ refer to the electrical conductivity and the magnetic permeability of the planetary core of radius r_0. In the case of the Earth, the decay time is of the order of 20 000 years, but it may vary between a few hundred and a million years for other examples of planetary dynamos.

The theory of homogeneous dynamos is based on Maxwell's equations for the magnetic flux density $\boldsymbol{B}$, the electric current density $\boldsymbol{j}$, and the electric field $\boldsymbol{E}$ in the MHD approximation in which the displacement current is neglected. This approximation is highly accurate as long as the fluid velocity is small compared with the velocity of light, which is certainly the case for all planetary applications. These equations together with Ohm's law for a moving conductor are given by

$$\nabla\cdot\boldsymbol{B}=0,\quad \frac{\partial}{\partial t}\boldsymbol{B}=-\nabla\times\boldsymbol{E} \qquad [1a]$$

$$\nabla\times\boldsymbol{B}=\mu\boldsymbol{j},\quad \boldsymbol{j}=\sigma(\boldsymbol{v}\times\boldsymbol{B}+\boldsymbol{E}) \qquad [1b]$$

where μ is the magnetic permeability of the fluid and σ is its electrical conductivity. These 'pre-Maxwell' equations have the property that they are invariant with respect to a Galilei transformation, that is, the equations remain unchanged in a new frame of reference moving with the constant velocity vector $\boldsymbol{V}$ relative to the original frame of reference. Indicating the variables of the new frame by a prime, we find

$$\boldsymbol{v}'=\boldsymbol{v}-\boldsymbol{V},\quad \frac{\partial}{\partial t'}=\frac{\partial}{\partial t}+\boldsymbol{V}\cdot\nabla \qquad [2a]$$

$$\boldsymbol{B}'=\boldsymbol{B},\quad \boldsymbol{E}'=\boldsymbol{E}+\boldsymbol{V}\times\boldsymbol{B},\quad \boldsymbol{j}'=\boldsymbol{j} \qquad [2b]$$

This invariance is the basis for the combination in MHD of eqns [1a] and [1b] with the equations of hydrodynamics in their usual nonrelativistic form. It is remarkable that this invariance does hold not only with respect to a Galilei transformation but also with respect to a transformation to a rotating frame of reference. In that case, $\boldsymbol{V}$ is replaced by $\boldsymbol{\Omega}\times\boldsymbol{r}$ in eqns [2a] and [2b], but when $\frac{\partial}{\partial t'}$ is operating on any vector $\boldsymbol{a}$, the term $-\boldsymbol{\Omega}\times\boldsymbol{a}$ must be added on the right-hand side, since even a constant vector field becomes time-dependent when seen from a rotating frame unless it is parallel to $\boldsymbol{\Omega}$.

Elimination of $\boldsymbol{E}$ and $\boldsymbol{j}$ from eqns [1a] and [1b] yields the equation of magnetic induction:

$$\nabla\times\left(\frac{1}{\sigma\mu}\nabla\times\boldsymbol{B}\right)=\frac{\partial}{\partial t}\boldsymbol{B}+\nabla\times(\boldsymbol{v}\times\boldsymbol{B}) \qquad [3]$$

which for a solenoidal velocity field υ and a constant magnetic diffusivity $\lambda=(\sigma\mu)^{-1}$ can be further simplified:

$$\left(\frac{\partial}{\partial t}+\boldsymbol{v}\cdot\nabla\right)\boldsymbol{B}-\lambda\nabla^2\boldsymbol{B}=\boldsymbol{B}\cdot\nabla\boldsymbol{v} \qquad [4]$$

This equation has the form of a heat equation with the magnetic field line stretching term on the right-hand side acting as a heat source. This interpretation is especially useful for the dynamo problem. In order that a magnetic field $\boldsymbol{B}$ may grow, the term on the right-hand side of eqn [4] must overcome the effect of the magnetic diffusion term on the left-hand side. Using a typical velocity U and a typical length scale d, the ratio of the two terms can be estimated by the magnetic Reynolds number Rm:

$$Ud/\lambda\equiv Rm \qquad [5]$$

Only when Rm is of the order one or larger may growing magnetic fields become possible.

In the following, we shall first consider the mathematical formulation of the problem of convection-driven dynamos in rotating spherical shells in a simple form in which only the physically most relevant parameters are taken into account. Before discussing dynamo solutions in Section 10.07.5, we shall briefly outline in Section 10.07.4 properties of convection

in the absence of a magnetic field. Applications to various planets and moons will be considered in Section 10.07.6 of this article, and some concluding remarks are given in Section 10.07.7.

10.07.3 Mathematical Formulation of the Problem of Spherical Dynamos

A sketch of the geometric configuration that will be considered is shown in Figure 1. For the equations describing convection-driven dynamos in the frame rotating with the angular velocity Ω, we use a standard formulation that has also been used for a dynamo benchmark (Christensen et al., 2001). But we use a different scaling and assume that a more general static state exists with the temperature distribution $T_S = T_0 - \beta d^2 r^2/2 + \Delta T \eta r^{-1}(1-\eta)^{-2}$ where η denotes the ratio of inner to outer radius of the spherical shell, d is its thickness, and rd is the distance from the center. ΔT is the temperature difference between the boundaries in the special case $\beta = 0$. In the case $\Delta T = 0$, the static temperature distribution T_S corresponds to that of a homogeneously heated sphere with the heat source density proportional to the parameter β. The gravity field is given by $g = -\gamma d\boldsymbol{r}$ where $\boldsymbol{r}$ is the dimensionless position vector with respect to the center of the sphere.

In addition to d, the time d^2/ν, the temperature $\nu^2/\boldsymbol{\gamma\alpha}\mathrm{d}^4$, and the magnetic flux density $\nu(\mu\varrho)^{1/2}/d$ are used as scales for the dimensionless description of the problem where ν denotes the kinematic viscosity of the fluid, κ its thermal diffusivity, and ϱ its density. The Boussinesq approximation is used in that ϱ is assumed to be constant except in the gravity term where its temperature dependence given by the constant $\alpha \equiv -(\mathrm{d}\varrho/\mathrm{d}T)/\varrho$ is taken into account. The dimensionless equations of motion, the heat equation for the deviation Θ of the temperature field from the static distribution, and the equation of magnetic induction thus assume the form

$$\nabla^2\boldsymbol{v} + \boldsymbol{B}\cdot\nabla\boldsymbol{B} + \boldsymbol{r}\Theta - \nabla\pi = P^{-1}(\partial_t\boldsymbol{v} + \boldsymbol{v}\cdot\nabla\boldsymbol{v}) + \tau\mathbf{k}\times\boldsymbol{v} \qquad [6a]$$

$$\nabla\cdot\boldsymbol{v} = 0 \qquad [6b]$$

$$\nabla^2\Theta + \left[R_\mathrm{i} + R_e\eta r^{-3}(1-\eta)^{-2}\right]\boldsymbol{r}\cdot\boldsymbol{v} = P(\partial_\mathrm{t} + \boldsymbol{v}\cdot\nabla)\Theta \qquad [6c)]$$

$$P_\mathrm{m}\left(\frac{\partial}{\partial t} + \boldsymbol{v}\cdot\nabla\right)\boldsymbol{B} - \nabla^2\boldsymbol{B} = P_\mathrm{m}\boldsymbol{B}\cdot\nabla\boldsymbol{v} \qquad [6d]$$

where $\mathbf{k}$ is the unit vector in the direction of the axis of rotation and where $\nabla\pi$ includes all terms that can be written as gradients. The Rayleigh numbers R_i and R_e, the Coriolis parameter τ, the Prandtl number P, and the magnetic Prandtl number P_m are defined by

$$R_\mathrm{i} = \frac{\alpha\gamma\beta d^6}{\nu\kappa}, \quad R_\mathrm{e} = \frac{\alpha\gamma\Delta T d^4}{\nu\kappa} \qquad [7a]$$

$$\tau = \frac{2\Omega d^2}{\nu}, \quad P = \frac{\nu}{\kappa}, \quad P_\mathrm{m} = \frac{\nu}{\lambda} \qquad [7b]$$

For simplicity, $R_e = 0$ will be assumed unless indicated otherwise. The notation $R \equiv R_i$ will thus be used in the following.

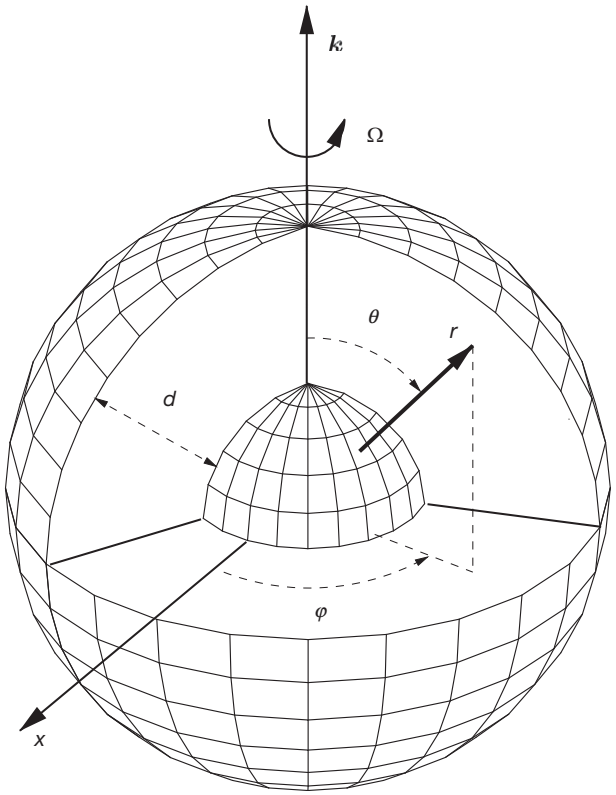


Figure 1 Geometric configuration of the problem. A part of the outer spherical surface is removed to expose the interior of the shell to which the conducting fluid is confined.

Instead of the Coriolis parameter τ, often its inverse, the Ekman number $E \equiv 2/\tau$, is used. The Prandtl number P has been added as an important parameter of the problem even though $P = 1$ is often assumed with the argument that all effective diffusivities are equal in turbulent media. The effective diffusivities of scalar and vector quantities in turbulent fluid flow differ in general, however, and large differences in the corresponding molecular diffusivities will not be erased entirely in the turbulent case.

Since the velocity field $\boldsymbol{v}$ and the magnetic flux density $\boldsymbol{B}$ are solenoidal vector fields, the general representation in terms of poloidal and toroidal components can be used:

$$\boldsymbol{v} = \nabla \times (\nabla \Phi \times \boldsymbol{r}) + \nabla \Psi \times \boldsymbol{r} \qquad [8a]$$

$$\boldsymbol{B} = \nabla \times (\nabla h \times \boldsymbol{r}) + \nabla g \times \boldsymbol{r} \qquad [8b]$$

By multiplying the $(\text{curl})^2$ and the curl of eqn [6a] by $\boldsymbol{r}$, we obtain two equations for Φ and Ψ:

$$\left[\left(\nabla^2 - \partial_t\right) L_2 + \tau \partial_\phi\right] \nabla^2 \Phi + \tau Q \Psi - L_2 \Theta$$

$$= -\boldsymbol{r} \cdot \nabla \times [\nabla \times (\boldsymbol{v} \cdot \nabla \boldsymbol{v} - \boldsymbol{B} \cdot \nabla \boldsymbol{B})] \qquad [9a]$$

$$\left[\left(\nabla^2 - \partial_t\right) L_2 + \tau \partial_\phi\right] \Psi - \tau Q \Phi = \boldsymbol{r} \cdot \nabla \times (\boldsymbol{v} \cdot \nabla \boldsymbol{v} - \boldsymbol{B} \cdot \nabla \boldsymbol{B}) \qquad [9b]$$

where ∂_t and ∂_ϕ denote the partial derivatives with respect to time t and with respect to the angle ϕ of a spherical system of coordinates r, θ, ϕ and where the operators L_2 and Q are defined by

$$L_2 \equiv -r^2 \nabla^2 + \partial_r (r^2 \partial_r)$$

$$Q \equiv r \cos\theta \nabla^2 - (L_2 + r\partial_r)(\cos\theta \partial_r - r^{-1} \sin\theta \partial_\theta)$$

The equations for h and g are obtained by multiplication of eqn [6d] and of its curl by $\boldsymbol{r}$:

$$\nabla^2 L_2 h = P_\mathrm{m} [\partial_t L_2 h - \boldsymbol{r} \cdot \nabla \times (\boldsymbol{v} \times \boldsymbol{B})] \qquad [10a]$$

$$\nabla^2 L_2 g = P_\mathrm{m} [\partial_t L_2 g - \boldsymbol{r} \cdot \nabla \times (\nabla \times (\boldsymbol{v} \times \boldsymbol{B}))] \qquad [10b]$$

Either rigid boundaries with fixed temperatures as in the benchmark case (Christensen et al., 2001),

$$\Phi = \partial_r (r\Phi) = \Psi = \Theta = 0 \quad \text{at } r = r_i \equiv \eta/(1-\eta) \quad \text{and at } r$$

$$= r_\mathrm{o} \equiv (1-\eta)^{-1} \qquad [11]$$

or stress-free boundaries with fixed temperatures,

$$\Phi = \partial_{rr}^2 \Phi = \partial_r (\Psi/r) = \Theta = 0 \quad \text{at } r = r_i$$

$$\equiv \eta/(1-\eta) \quad \text{and at } r = r_o = (1-\eta)^{-1} \qquad [12]$$

are often used. The latter boundary conditions are assumed in the following since they allow to cover numerically a larger region of the parameter space. The case $\eta = 0.4$ will be considered unless indicated otherwise. It provides a good compromise for the study of both the regions inside and outside the tangent cylinder. The latter is the cylindrical surface touching the inner spherical boundary at its equator. For the magnetic field, electrically insulating boundaries are used such that the poloidal function h must be matched to the function $h^{(e)}$ that describes the potential fields outside the fluid shell:

$$g = h - h^{(e)} = \partial_r \left(h - h^{(e)}\right) = 0 \quad \text{at } r = r_i \quad \text{and } r = r_o \qquad [13]$$

Computations for the case of an inner boundary with no-slip conditions and an electrical conductivity equal to that of the fluid have also been done. The numerical integration of eqns [2a] and [2b] together with boundary conditions [4] proceeds with the pseudospectral method as described by Tilgner and Busse (1997) and Tilgner (1999), which is based on an expansion of all dependent variables in spherical harmonics for the θ, ϕ dependences, that is,

$$\Phi = \sum_{l,m} V_l^m(r,t) P_l^m(\cos\theta) \exp\{im\phi\} \qquad [14]$$

and analogous expressions for the other variables, Ψ, Θ, h, and g. P_l^m denotes the associated Legendre functions. For the r-dependence, expansions in Chebyshev polynomials are used.

For the computations to be reported in the following, a minimum of 33 collocation points in the radial direction and spherical harmonics up to the order 64 have been used. But in many cases, the resolution was increased to 49 collocation points and spherical harmonics up to the order 96 or 128.

It should be emphasized that the static state $\boldsymbol{v} = \boldsymbol{B} = \Theta = 0$ represents a solution of eqns [6a]–[6d] for all values of the Rayleigh numbers R_i and R_e, but this solution is unstable except for low or negative values of the latter parameters. Similarly, there exist solutions with $\boldsymbol{B} = 0$, but $\boldsymbol{v} \neq 0$, $\Theta \neq 0$, for sufficiently large values of either R_i or R_e or both, but, again, these solutions are unstable for sufficiently large values of P_m with respect to disturbances with $\boldsymbol{B} \neq 0$. Dynamo solutions as all solutions with $\boldsymbol{B} \neq 0$ are called, are thus removed by at least two bifurcations from the basic static solution of the problem.

Finally, we present in **Table 1** a list of the most important parameters used in the following sections.

10.07.4 Convection in Rotating Spherical Shells

For an introduction to the problem of convection in spherical shells, we refer to the review of Busse (2002a). Convection tends to set in first outside the tangent cylinder in the form of thermal Rossby waves for which the Coriolis force is balanced almost entirely by the pressure gradient. The model of the rotating cylindrical annulus has been especially useful for

Table 1 Important dimensionless parameters

E	Average density of kinetic energy	[18a] and [18b]
E_p, E_t	Energy densities of poloidal and toroidal components of motion	[18a] and [18b]
η	Radius ratio of spherical fluid shell	
Λ	Elsasser number	[19]
M	Average magnetic energy density	[19]
M_p, M_t	Energy densities of poloidal and toroidal components of the magnetic field	[19]
Nu	Nusselt number	[16]
P	Prandtl number	[7b]
P_m	Magnetic Prandtl number	[7b]
R	Rayleigh number	[7a]
R_c	Critical value of R for the onset of convection	
Rm	Magnetic Reynolds number $= P_\mathrm{m}\sqrt{2E}$	
τ	Coriolis parameter	[7b]

the understanding of this type of convection. A rough idea of the dependence of the critical Rayleigh number R_c for the onset of convection on the parameters of the problem can be gained from the expressions derived from the annulus model (Busse, 1970):

$$R_c = 3\left(\frac{P\tau}{1+P}\right)^{\frac{4}{3}}(\tan\theta_m)^{\frac{8}{3}} r_m^{-\frac{1}{3}} 2^{-\frac{2}{3}} \qquad [15a]$$

$$m_c = \left(\frac{P\tau}{1+P}\right)^{\frac{1}{3}}(r_m\tan\theta_m)^{\frac{2}{3}} 2^{-\frac{1}{6}} \qquad [15b]$$

$$\omega_c = \left(\frac{\tau^2}{(1+P)^2 P}\right)^{\frac{1}{3}} 2^{-\frac{5}{6}}(\tan^2\theta_m/r_m)^{\frac{2}{3}} \qquad [15c]$$

where r_m refers to the mean radius of the fluid shell, $r_m=(r_i+r_o)/2$, and θ_m refers to the corresponding colatitude, $\theta_m=\arcsin(r_m(1-\eta))$. The azimuthal wavenumber of the preferred mode is denoted by m_c, and the corresponding angular velocity of the drift of the convection columns in the prograde direction is given by ω_c/m_c.

In Figure 2, expressions [15a] and [15c] are compared with accurate numerical values that indicate that the general trend is well represented by expressions [15a] and [15c]. The same property holds for m_c. For a rigorous asymptotic analysis in the case $P=1$ including the radial dependence, we refer to Jones et al. (2000).

It is evident from Figure 2 that the agreement between expressions [15a]–[15c] and the numerical values deteriorates as low values of P are approached. This behavior is caused by the fact that instead of the thermal Rossby wave mode, the inertial mode of convection becomes preferred at onset for sufficiently low Prandtl numbers. It is characterized by convection cells attached to the equatorial region of the outer boundary not unlike the 'banana cells' seen in the narrow gap experiment of Figure 3. The equatorially attached convection does indeed represent an inertial wave modified by the effects of viscous dissipation and thermal buoyancy. An analytic description of this type of convection can thus be attained through the introduction of viscous friction and buoyancy as perturbations, as has been done by Zhang (1994) and Busse and Simitev (2004) for stress-free boundaries and by Zhang (1995) for no-slip boundaries. According to Ardes et al. (1997), equatorially attached convection is preferred at onset for $\tau<\tau_1$ where τ_1 increases in proportion to $P^{-1/2}$.

A third form of convection is realized in the polar regions of the shell that comprise the two fluid domains inside the tangent cylinder. Since gravity and rotation vectors are nearly parallel in these regions (unless values of η close to unity are used), convection resembles the kind realized in a horizontal layer heated from below and rotating about a vertical axis. Because the Coriolis force cannot be balanced by the pressure gradient in this case, the onset of convection is delayed to higher values of R where convection outside the tangent cylinder has reached already high amplitudes. In the case of $\eta=0.4$, the onset of convection in the polar regions typically occurs at the Rayleigh numbers that exceed the critical values R_c for the onset of convection outside the tangent cylinder by a factor of the order 4. Except for the case of very low Prandtl numbers, the retrograde differential rotation in the polar regions generated by convection outside the tangent cylinder tends to facilitate polar convection by reducing the rotational constraint. A tendency towards an alignment of convection rolls with the north–south direction (Busse and Cuong, 1977) can be noticed, but this property is superseded by the instabilities of the Küppers–Lortz type (for an experimental demonstration, see Busse and Heikes, 1980) and by interactions with turbulent convection outside the tangent cylinder.

Typical features of low and high Prandtl number convection are illustrated in Figure 4. The columnar nature of

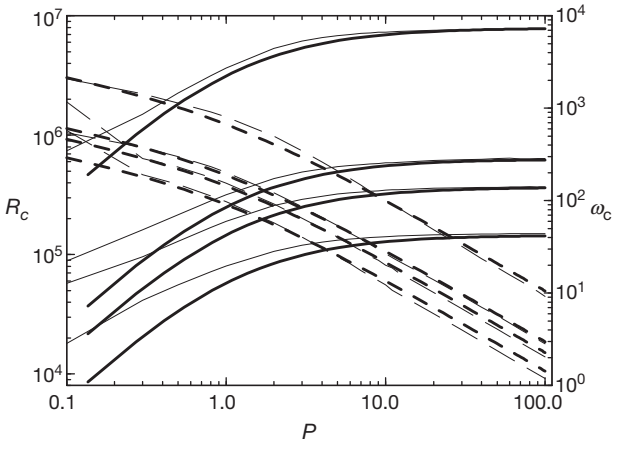


Figure 2 Critical Rayleigh number R_c (thin solid lines) and frequency ω_c (right ordinate, thin dashed lines) as a function of the Prandtl number P in the case $\eta=0.4$ for the Coriolis numbers $\tau=5\times10^3$, 10^4, 1.5×10^4, and 10^5 (from bottom to top). The thick lines correspond to expressions [15a] and [15c].

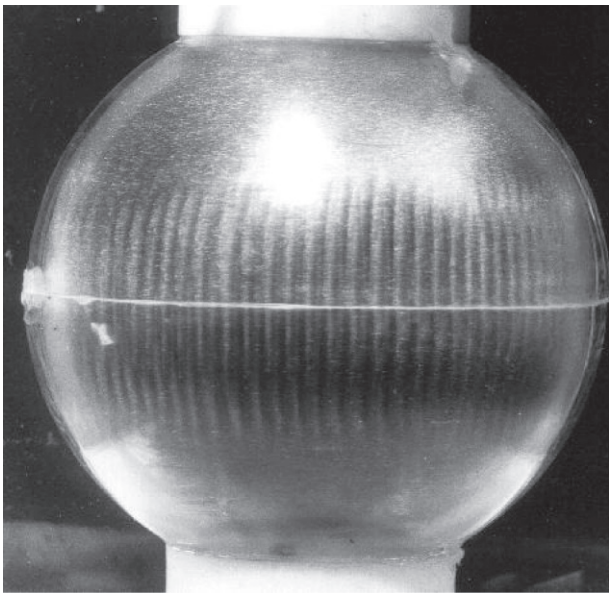


Figure 3 Banana cells in a thin rotating spherical fluid shell cooled from within. Convection driven by centrifugal buoyancy is made visible by a suspension of tiny flakes that become aligned with the shear. Reproduced from Busse FH and Carrigan CR (1976) Laboratory simulation of thermal convection in rotating planets and stars. *Science* 191: 81–83.

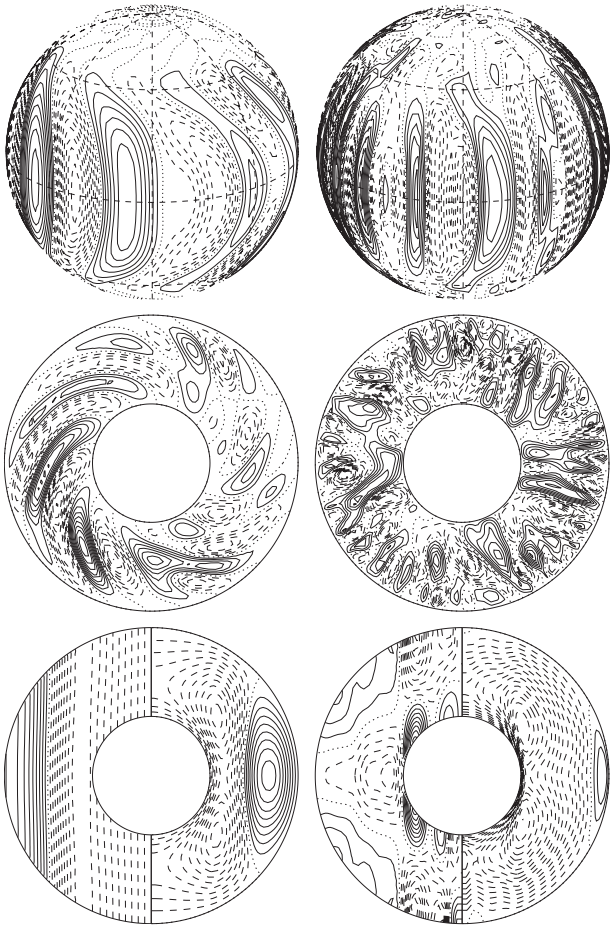


Figure 4 Convection in rotating spherical fluid shells in the cases $\tau = 10^4$, $R = 4 \times 10^5$, and $P = 1$ (left column) and $\tau = 5 \times 10^3$, $R = 8 \times 10^5$, and $P = 20$ (right column). Lines of constant u_r in the middle spherical surface, $r = r_i + 0.5$, are shown in the upper row. The plots of the middle row show streamlines, $r\partial\Phi/\partial\varphi = \text{const.}$, in the equatorial plane. The lowermost plots indicate lines of constant mean azimuthal velocity $\overline{u}_\varphi$ in the left halves and isotherms of $\overline{\Theta}$ in the right halves.

convection does not vary much with P, as is evident from the top two plots of the figure. At Prandtl numbers of the order unity or less, but not in the case of inertial convection, the convection columns tend to spiral away from the axis and thereby create a Reynolds stress that drives a strong geostrophic differential rotation, as shown in the bottom left plot of the figure. This differential rotation in turn promotes the tilt and a feedback loop is thus created. The generation of differential rotation by convection in rotating spherical shells was first demonstrated by Busse (1970). In this paper, a simple analytic model is described (see also Busse, 2002a). At high values of P, the Reynolds stress becomes negligible, and no significant tilt of the convection columns is apparent. In this case, the differential rotation is generated in the form of a thermal wind caused by the latitudinal gradient of the axisymmetric component of Θ.

Among the properties of convection at finite amplitudes, the heat transport is the most important one. Customarily, its efficiency is measured by the Nusselt number that is defined as the heat transport in the presence of convection divided by the

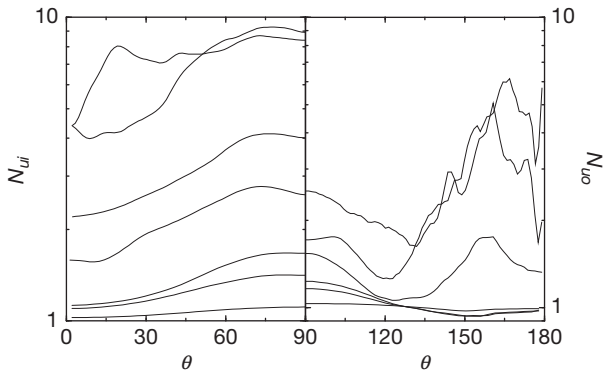


Figure 5 The time-averaged and azimuthally averaged local Nusselt numbers at the inner spherical boundary *Nu_i* (left plot) and at the outer spherical boundary *Nu_o* (right plot) as functions of the colatitude θ for $P = 0.5$, $\tau = 1.5 \times 10^4$, and $R = 4$, 8, 10, 15, 20, 30, 25×10^5 (from bottom to top at 90°).

heat transport in the absence of motion. In the case of the spherical fluid shell, two Nusselt numbers can be defined measuring the efficiency of convection at the inner and the outer boundary:

$$Nu_i = 1 - \frac{P}{r_i}\frac{d\overline{\overline{\Theta}}}{dr}\bigg|_{r=r_i} \qquad Nu_o = 1 - \frac{P}{r_o}\frac{d\overline{\overline{\Theta}}}{dr}\bigg|_{r=r_o} \qquad [16]$$

where the double bar indicates the average over the spherical surface. In addition, local Nusselt numbers

$$Nu_i(\theta) = 1 - \frac{P}{r_i}\frac{d\overline{\Theta}}{dr}\bigg|_{r=r_i} \qquad Nu_o(\theta) = 1 - \frac{P}{r_o}\frac{d\overline{\Theta}}{dr}\bigg|_{r=r_o} \qquad [17]$$

are of interest where only the azimuthal average is applied, as indicated by the single bar, instead of the average over the entire spherical surface. Examples of such measures of the dependence of the heat transport on latitude are shown in Figure 5. This figure demonstrates that at low supercritical Rayleigh numbers, the heat transport occurs primarily across the equatorial region, but as R increases, the heat transport in the polar regions takes off and soon exceeds that at low latitudes. This process is especially evident at the outer boundary. In the polar regions, convection is better adjusted for carrying heat from the lower boundary to the upper one, and it is known from computations of the convective heat transport in horizontal layers rotating about a vertical axis that the value of Nu may exceed the value in a nonrotating layer at a given value of R in spite of the higher critical Rayleigh number in the former case (Somerville and Lipps, 1973; Stevens et al., 2011).

Outside the tangent cylinder, the convective heat transport encounters unfavorable conditions in that the cylindrical form of the convection eddies is not well adjusted to the spherical boundaries. This handicap is partly overcome through the onset of time dependence in the form of vacillations in which the convection columns expand and contract in the radial direction or vary in amplitude.

At Prandtl numbers of the order unity and less, another effect restricts the heat transport. The shear of the geostrophic differential rotation created by the Reynolds stresses of the convection columns severely inhibits the heat transport. To

illustrate this effect, we have plotted in Figure 6 in addition to the Nusselt numbers the averages of the kinetic energy densities of the various components of the convection flow, which are defined by

$$\overline{E}_{\rm p}=\frac{1}{2}\left\langle\left|\nabla\times\left(\nabla\overline{\Phi}\times r\right)\right|^2\right\rangle,\quad \overline{E}_{\rm t}=\frac{1}{2}\left\langle\left|\nabla\overline{\Psi}\times r\right|^2\right\rangle \qquad [18a]$$

$$E_{\rm p}=\frac{1}{2}\left\langle\left|\nabla\times\left(\nabla\overset{\vee}{\Phi}\times r\right)\right|^2\right\rangle,\quad \overset{\vee}{E}_{\rm t}=\frac{1}{2}\left\langle\left|\nabla\overset{\vee}{\Phi}\times r\right|^2\right\rangle \qquad [18b]$$

where the angular brackets indicate the average over the fluid shell and $\overline{\Phi}$ refers to the azimuthally averaged component of Φ, while $\overset{\vee}{\Phi}$ is defined by $\overset{\vee}{\Phi}=\Phi-\overline{\Phi}$. Analogous definitions hold for the energy densities of the magnetic field where E, Φ, and Ψ are replaced by M, h, and g, respectively. The total magnetic energy density M averaged over the fluid shell is thus given by $M=\overline{M}_{\rm p}+\overline{M}_{\rm t}+\overset{\vee}{M}_{\rm p}+\overset{\vee}{M}_{\rm t}$.

Figure 6 is instructive in that it demonstrates convection both in the presence and in the absence of its dynamo-generated magnetic field. As is evident from the right part of Figure 6 with $M\sim 0$, relaxation oscillations occur in which convection sets in nearly periodically for short episodes once the energy $\overline{E}_{\rm t}$ of the differential rotation has decayed to a sufficiently low amplitude. But as soon as convection grows in amplitude, the differential rotation grows as well and shears off the convection columns. After convection has stopped, the differential rotation decays on the viscous timescale until the process repeats itself.

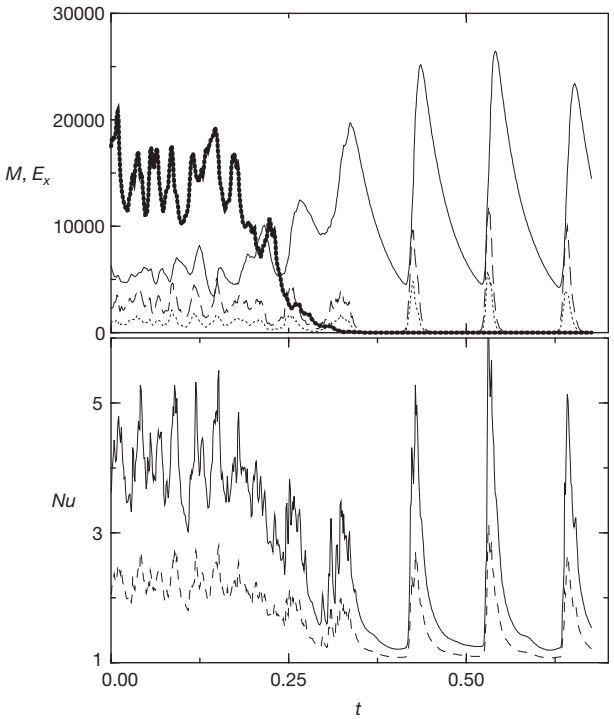


Figure 6 *Decay of a dynamo for* $P=0.5$, $\tau=1.5\times10^4$, $R=1.2\times10^6$, and $P_{\rm m}=0.5$. The total magnetic energy density M (thick dotted line, multiplied by a factor of 5), the kinetic energy densities $\overline{E}_{\rm t}$ (solid line), $\check{E}_{\rm p}$ (dotted line), $\check{E}_{\rm t}$ (dashed line), and the Nusselt numbers Nu_i (solid line) at the inner spherical boundary and Nu_o (dashed line) at the outer spherical boundary (lower plot) are plotted as functions of time t.

As long as the magnetic field is present, $\overline{E}_{\rm t}$ is suppressed owing to the action of the Lorentz force and high Nusselt numbers are obtained. In a nutshell, Figure 6 thus demonstrates why planets and stars require dynamos. Without them, they would not get rid of their heat in an efficient way due to the constraints of rotation.

It is worth pointing out that the dynamo-generated magnetic field acts in a fashion quite different from that of an imposed homogeneous field, which typically counteracts the effects of rotation and tends to minimize the critical value of the Rayleigh number when the Elsasser number,

$$\Lambda=\frac{2MP_{\rm m}}{\tau} \qquad [19]$$

assumes the value 1 in the case of a plane layer (Chandrasekhar, 1961) or values of the same order in the case of a sphere (Fearn, 1979) or in the related annulus problem (Busse, 1983).

10.07.5 Convection-Driven Dynamos

It appears that dynamos are generated by convection in rotating spherical shells for all parameter values as long as the magnetic Reynolds number is of the order 50 or higher and the fluid is not too turbulent. Since Rm can be defined by $P_{\rm m}\sqrt{2E}$ where the kinetic energy density, $E=\overline{E}_{\rm p}+\overline{E}_{\rm t}+\overset{\vee}{E}_{\rm t}+\overset{\vee}{E}_{\rm t}$, increases with R, increasing values of the Rayleigh number are required for dynamos as $P_{\rm m}$ decreases. In planetary cores, the latter parameter may assume values as low as 10^{-5} or 10^{-6}, but numerical simulations have achieved so far only values somewhat below 10^{-1}. The trend of increasing R with decreasing $P_{\rm m}$ is evident in Figure 7 where results are shown for two different values of τ and P. It is also evident from the results in the upper left corner of the figure that an increasing R may be detrimental for dynamo action. This is a typical property of marginal dynamos at the boundary of dynamos in the parameter space

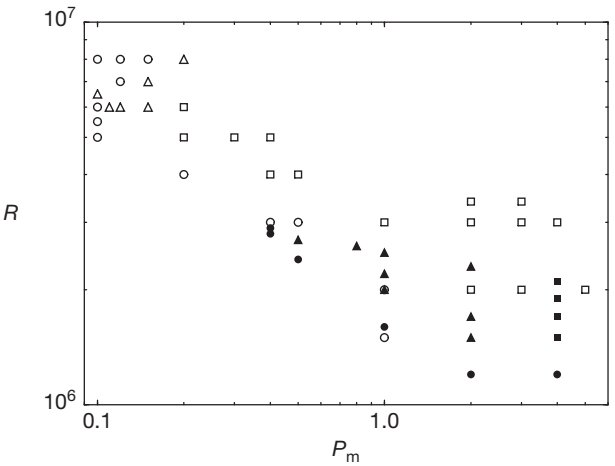


Figure 7 Convection-driven dynamos as a function of the Rayleigh number R and the magnetic Prandtl number $P_{\rm m}$ for $P=0.1$, $\tau=10^5$ (empty symbols) and for $P=1$, $\tau=3\times10^4$ (filled symbols). The symbols indicate chaotic dipolar (squares), hemispheric (triangles), and decaying dynamos (circles).

(Christensen et al., 1999; Simitev and Busse, 2005). Since convection at the relevant values of R is chaotic, this property also holds for the generated magnetic field. The magnetic energy must be finite near the boundary of dynamos in the parameter space because a sufficient amplitude of the fluctuating components of the convection flow can be obtained only when the magnetic field is strong enough to suppress the differential rotation. This property appears to hold even at high Prandtl numbers where differential rotation occurs only in the form of a relatively weak thermal wind.

Even in its chaotic state, convection at high values of τ exhibits a strong symmetry with respect to the equatorial plane at high values of τ, at least as long as convection in the polar regions is still weak. Because of this symmetry, either magnetic fields can be generated with the same symmetry as the convection flow, that is, h and g are antisymmetric with respect to equatorial plane in which case one speaks of a quadrupolar dynamo, or the magnetic field exhibits the opposite symmetry with symmetrical functions h and g in which case one speaks of a dipolar dynamo. Of special interest are hemispheric dynamos (Grote and Busse, 2001) in which case the fields of dipolar and quadrupolar symmetry have nearly the same amplitude such that they cancel each other either in the Northern or in the Southern Hemisphere. Examples of typical dynamos with different symmetries are shown in Figure 8. Here, it is also evident that low P and high P dynamos differ in the structure of their mean toroidal fields. While their mean poloidal dipolar fields exhibit hardly any difference, the high P dynamo is characterized by strong polar azimuthal flux tubes that are missing in the low P case. The reason for this difference is that the radial gradient of the differential rotation in the polar regions is much stronger for high P than for low P.

As the Rayleigh number increases and polar convection becomes stronger, the convection flow loses some of its equatorial symmetry and the magnetic field can no longer easily be classified. Usually, the dipolar component becomes more dominant in cases of dynamos that started as either quadrupolar or hemispheric dynamos at lower values of R.

It should be emphasized that quadrupolar and hemispheric dynamos are usually oscillatory in that new azimuthal flux tubes of alternating sign emerge in the equatorial plane and move towards higher latitudes, while old flux tubes are dissipated in the polar region. There also exist dipolar dynamos that exhibit the same type of oscillations. They are typically found in the parameter space near the region where hemispheric dynamos occur. To illustrate the dynamo oscillations, additional plots separated at equal distances in time before the corresponding plot in the lower three rows of Figure 8 are shown in Figure 9. The four plots in each case cover approximately half a period of oscillation. It must be realized, of course, that the oscillations are not strictly periodic since they occur in a turbulent system. In this respect, they resemble the solar cycle with its 22-year period. Because the solar dynamo is believed to operate at the bottom of the solar convection zone, the propagation of the dynamo wave is towards low latitudes on the sun. A remarkable feature of the dipole oscillation in Figure 9 is that the polar flux tubes hardly change in time. The oscillation appears to be confined to the region outside the tangent cylinder. At a lower value of P_m, the strong polar flux tubes even inhibit the oscillation of the mean poloidal field as

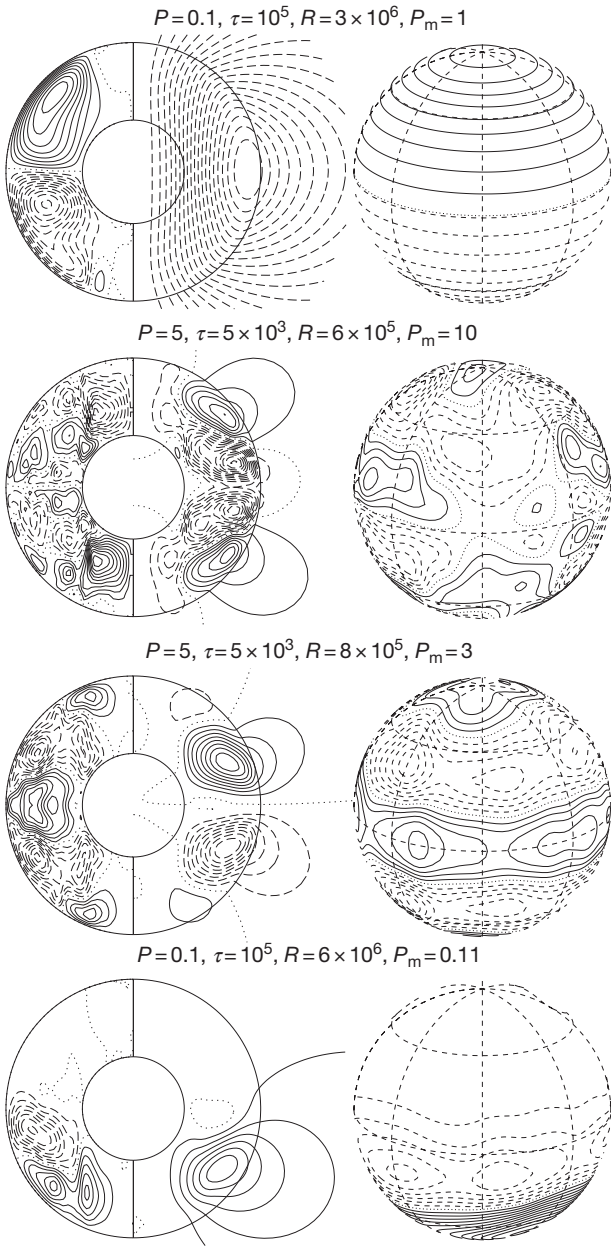

$P=0.1,\ \tau=10^5,\ R=3\times10^6,\ P_m=1$
$P=5,\ \tau=5\times10^3,\ R=6\times10^5,\ P_m=10$
$P=5,\ \tau=5\times10^3,\ R=8\times10^5,\ P_m=3$
$P=0.1,\ \tau=10^5,\ R=6\times10^6,\ P_m=0.11$

Figure 8 Various types of dynamo symmetry. The left column shows lines of $\overline{B}_\varphi = \text{const.}$ (left half) and $r\sin\theta\partial_\theta\overline{h} = \text{const.}$ (right half). The right column shows surfaces of $B_r = \text{const.}$ at $r = r_o + 0.7$.

shown in Figure 10. This case has been called the 'invisibly' oscillating dynamo since at some distance from the boundary of the spherical fluid shell, the oscillation of the dynamo can hardly be noticed.

The relatively simple magnetic fields displayed in Figures 8–10 should not be regarded as representative for high values of the Rayleigh number when R exceeds its critical value R_c by an order of magnitude or more. In that case, the equatorial symmetry has nearly disappeared owing to the dominant convection in the polar regions, and nonaxisymmetric components of the field tend to exceed axisymmetric ones as shown, for example, in Figure 11. The strong growth with

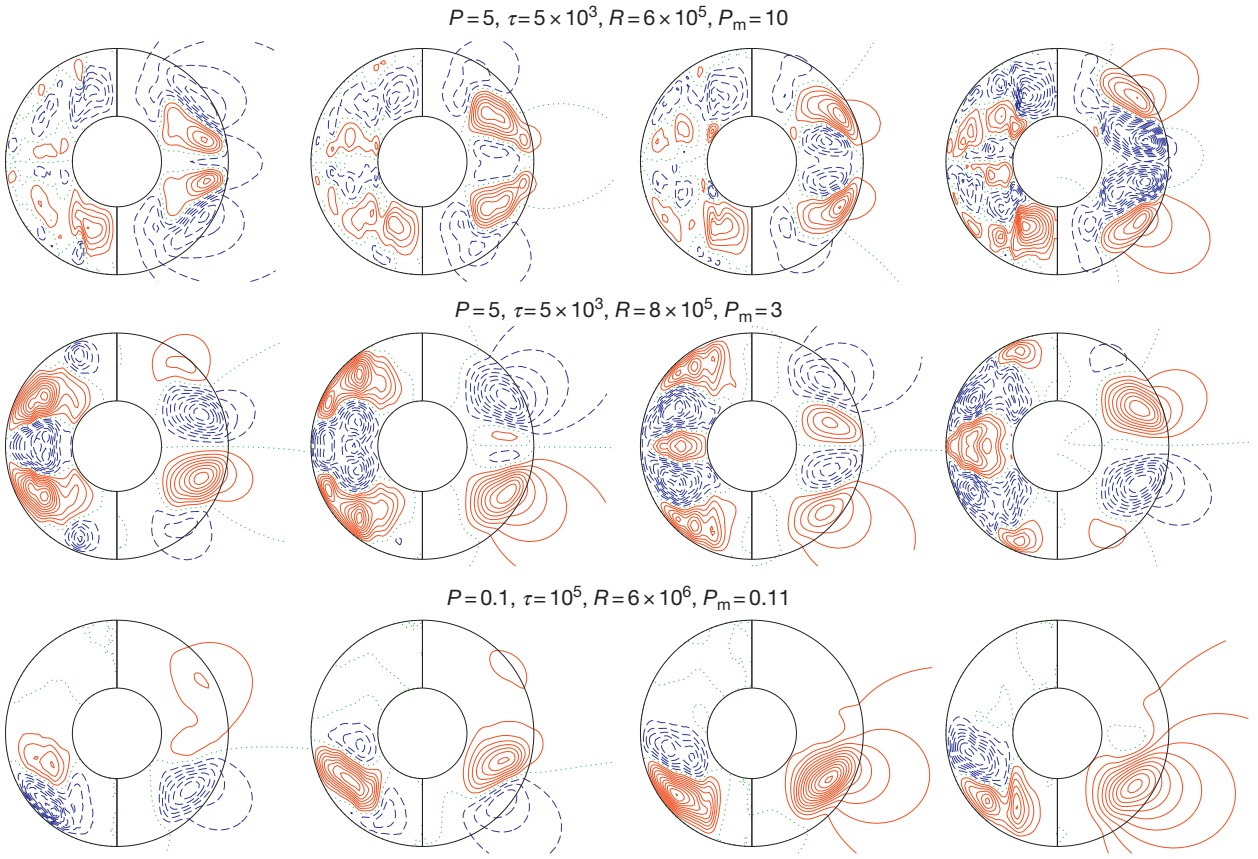


Figure 9 Plots for the last three cases of **Figure 8** at the times $n\cdot\Delta t$, $n=3, 2, 1, 0$ (left to right), before the times of **Figure 8** where $\Delta t=0.035$, 0.04, and 0.0025 (from top to bottom) has been used.

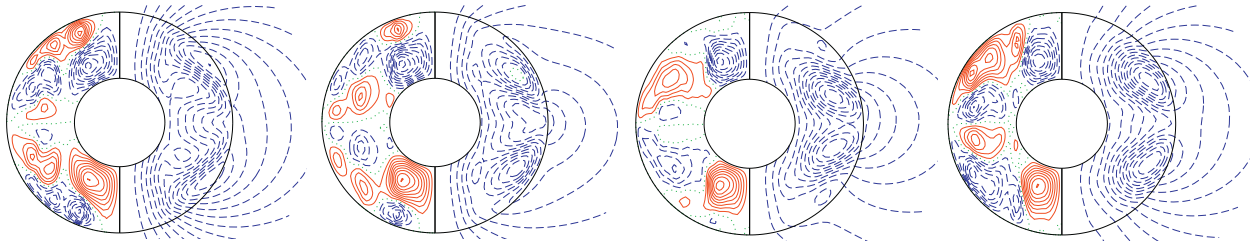


Figure 10 An 'invisibly' oscillating dynamo with $\Delta t=0.04$ in the case of $P=P_m=5$, $\tau=5\times10^3$, and $R=6\times10^5$. The plots show lines of constant $\overline{B}_\varphi$ in their left halves and meridional field lines, $r\sin\theta\partial_\theta\bar{h}=$ const., in their right halves. Δt represents about a quarter of a cycle.

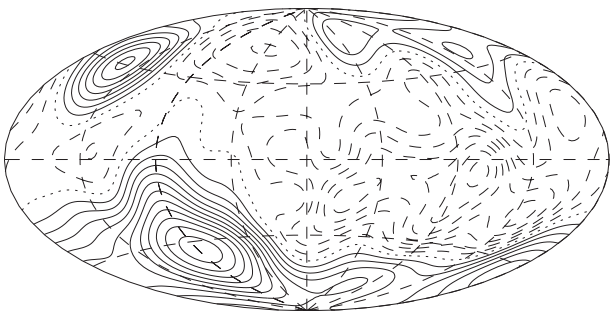


Figure 11 Lines of constant B_r at $r=r_o+1$ for the dynamo with $R=2\times10^6$, $\tau=5\times10^3$, and $P=P_m=1$.

increasing Rm of the nonaxisymmetric magnetic flux is also evident in **Figure 12** when energies corresponding to different values of P_m are compared.

An important question is the average strength of the magnetic field generated by the dynamo process in dependence on the parameters. Two concepts are often used in attempts to answer this question. In astrophysical situations, the equipartition between magnetic energy and kinetic energy is used as a guide. While such a balance may hold locally as, for example, in the neighborhood of sunspots, it is not likely to be applicable to global planetary magnetic fields. In the Earth's core, for example, the magnetic energy density exceeds the kinetic one by a factor of the order of 10^3.

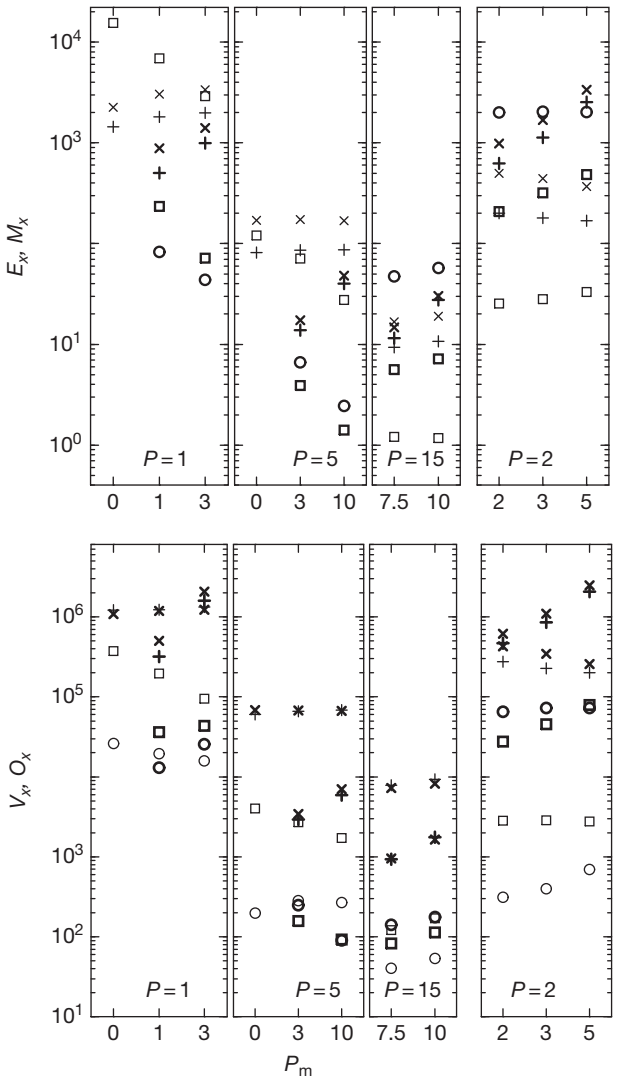


Figure 12 Kinetic E_x and magnetic M_x energy densities (upper part) and viscous V_x and Ohmic O_x dissipations (lower part) as functions of P_m for convection-driven dynamos for $\tau=5\times10^3$, $R=10^6$ (left three columns), $\tau=3\times10^4$, and $R=3.5\times10^6$ (right column) and the Prandtl number as indicated in the boxes. The highest values of the Elsasser number for the cases $P=1$, 5, 15, and 2 are $\Lambda=3.02$, 0.37, 0.49, and 1.12, respectively. The components $\overline{X}_p, \overline{X}_t, \widetilde{X}_p, \widetilde{X}_t$ (where X = E, M, V, O) are represented by circles, squares, plus-signs, and crosses, respectively. Kinetic energy densities and viscous dissipations are shown with light symbols; magnetic energy densities and Ohmic dissipations are shown with heavy symbols. The magnetic Prandtl numbers $P_m=0$ correspond to nonmagnetic convection.

A second concept is based on the property that an Elsasser number Λ of the order unity appears to correspond to optimal conditions for convection in rotating systems in the presence of an applied nearly uniform magnetic field. While this idea may be useful as a first rough estimate, the results of numerical dynamo simulations do not really support this concept. The examples shown in Figure 12 exhibit Λ-values differing by an order of magnitude, and even larger variations have been reported by Simitev and Busse (2005) and by Christensen and Aubert (2006). For a critical assessment of the use of the Elsasser number, see Soderlund et al. (2012). It should also be kept in mind that at the same conditions, two distinct types of dynamos may exist that exhibit vastly different mean dipolar fields (Simitev and Busse, 2009, 2012).

The large number of computationally generated convection-driven dynamos has allowed Christensen and Aubert (2006) to obtain a scaling relationship for the strength B of the magnetic field that no longer depends on the rate of rotation Ω but instead depends only on the available power for driving the dynamo. They analyze their data in terms of parameters such as the Rayleigh number, Nusselt number, and the dimensionless buoyancy flux $\hat{Q}$ that are defined through quantities that no longer involve molecular diffusivities. This procedure requires the presence of rotation and cannot be achieved in a non-rotating system. It finally leads to the scaling relationship

$$B^2 \propto \rho r_o^{4/3} \hat{Q}^{2/3} \qquad [20]$$

for the magnetic energy density in the dynamo region that fits the available numerical data for sufficiently high rates of rotation quite well. For a more detailed discussion of scaling laws, see Jones (2011).

The brief introduction of this section to convection-driven dynamos in rotating spherical fluid shells can only give a vague impression of the potential of numerical simulations for the understanding of planetary magnetism. Many more examples of such simulations can be found in the literature. Usually, they have been motivated by applications to the geodynamo, and $P=1$ is assumed in most cases for simplicity. For some recent systematic investigations, we refer to Christensen et al. (1999), Grote et al. (2000), Jones (2000), Grote and Busse (2001), Kutzner and Christensen (2000, 2002), Busse et al. (2003), Simitev and Busse (2005), Aubert et al. (2009), and other papers referred to therein.

10.07.6 Applications to Planetary Dynamos

10.07.6.1 General Considerations

Although numerical simulations of the kind discussed in the preceding sections are still rather removed from the parameter regime relevant to planetary interiors, numerous dynamo models with modified boundary conditions or including regions of stably stratified density have been applied with some success to planetary situations. Properties such as compressibility and other deviations from the Boussinesq approximation that so far have been considered only in special models applied to the Earth's core (Glatzmaier and Roberts, 1996) or to the sun (Brun et al., 2004) are beginning to be taken into account for dynamo models of the giant planets. In this respect, the benchmark for anelastic convection and dynamos (Jones et al., 2011) is of interest.

Many essential parameter values of planetary cores are not sufficiently well known, however, to provide a basis for the development of more specific dynamo models. Much of the future progress of the field will thus depend on the mutual constraints derived from observational evidence and from theoretical conclusions in order to arrive at a better understanding of the workings of planetary dynamos.

In Table 2, some properties related to planetary magnetism have been listed. The dipole moments of the planets have been given as multiples of $G \cdot R_p^3$ where R_p denotes the mean radius of the planet or satellite. Thus, the numerical value indicates the field strength in *gauss* in the equatorial plane of the dipole at the distance R_p from its center. We have included Mars in the table, although it does have an active dynamo at the present time. The strong magnetization of the Martian crust indicates, however, that a strong field dynamo must have operated in the early history of the planet (see also **Chapter 10.06**).

The most important question of the theory of planetary magnetism is the dependence of the observed field strength on the properties of the planet. We have already discussed in the historical introduction various proposals for such dependences based on ad hoc assumptions, but these have been abandoned by and large. Only the concept of an Elsasser number Λ of the order unity is still frequently used. Stevenson (2003) showed that a value $\Lambda = 0.3$ fits most of the planets with a global magnetic field quite well, but in cases such as Mercury and Saturn, only a value of the order 10^{-2} can be estimated for Λ. In order to save the concept, different dynamo regimes must be assumed. Equilibration balances for weaker magnetic fields have been proposed as, for instance, in the case when the strength of convection is not sufficient to attain the $\Lambda \sim 1$ balance. For details, see Stevenson (1984).

An important balance often invoked in discussions of planetary dynamos (see, e.g., Jones, 2003; Stevenson, 1979) is the MAC balance (Braginsky, 1967) where it is assumed that Lorentz force, buoyancy force, and pressure gradient are all of the same order as the Coriolis force, while viscous friction and the momentum advection term are regarded as negligible in the equations of motion. The neglect of the latter term is well justified for high Prandtl numbers, but it is doubtful for values of P of the order unity or less. In spite of their smallness in comparison with the Coriolis force, the divergence of the Reynolds stress can generate the most easily excitable mode of a rotating fluid, namely, the geostrophic differential rotation, which cannot be driven by the Coriolis force, the pressure gradient, or buoyancy. Of course, an excitation by the Lorentz force is possible in principle. The latter force, however, usually inhibits the geostrophic differential rotation. Since the mean azimuthal component of the magnetic field is typically created through the shear of the differential rotation, the Lorentz force opposes the latter according to Lenz's rule.

The prediction [20] of Christensen and Aubert (2006) for the strength of the magnetic field in the dynamo region fits the cases of the Earth and Jupiter quite well but yields the result that Mercury's field could not be generated by a buoyancy-driven dynamo since the magnetic Reynolds number would be too small. In the cases of the outer planets, relevant parameters are not sufficiently well known to draw definitive conclusions. The negative result in the case of Mercury may be misleading, however, since the observed field may not be representative for the dynamo-generated field inside the planet as will be discussed in **Section 10.07.6.2**.

Finally, we like to draw attention to what has been called Stevenson's paradox, that is, that a too high electrical conductivity is detrimental for dynamo action in planetary cores. According to the Wiedemann–Franz law of condensed matter physics, the thermal conductivity k of a metal is proportional to its electrical conductivity times the temperature, $k = \sigma TL$, where L has approximately the same value for all metals. This is due to the fact that in metals, electrons dominate the transports of both heat and charge. As a consequence, there is an upper bound on the electrical conductivity for which a dynamo driven by thermal convection is possible because a high value of σ implies a low temperature change with increasing depth needed to carry the heat from the interior of the planetary core. But as soon as the temperature change with pressure falls below its isentropic value, thermal convection disappears and a stably stratified environment is obtained, which makes it difficult to drive a dynamo by other types of motions such as compositional convection. Ignoring the latter possibility, Stevenson (2003) had sketched the diagram shown in **Figure 13** where the marginal nature of dynamos in terrestrial and in the 'icy' planets is indicated.

Related to Stevenson's paradox is the fact that regions of high electrical conductivity are not the most likely places where dynamos operate in the major planets. Magnetic field generation in deeper regions of Jupiter and Saturn would lead to an Ohmic dissipation exceeding the luminosity of these planets, as has been emphasized by Liu et al. (2008). It thus appears

Table 2 Planetary parameter values

Planet (satellite)	*Planetary radius*, R_p ($\times 10^6$ m)	*Dipole moment* ($\times 10^{-4}$ T $\times R_p^3$)	*Core radius* ($\times 10^6$ m)	*Angular rate of rotation*, Ω ($\times 10^{-5}$ s^{-1})	*Magnetic diffusivity* (m^2 s^{-1})
Mercury	2.439	0.0025	2.0	0.124	~2
Earth	6.371	0.31	3.48	7.29	2
Mars	3.389	3[a]	~1.5	7.08	~2
Jupiter	69.95	4.3	60[b]	17.6	100[b]
Saturn	58.30	0.21	35[b]	16.2	100[b]
Uranus	25.36	0.23	17[b]	10.1	100[b]
Neptune	24.62	0.14	18[b]	10.8	100[b]
Ganymede	2.63	0.0072	0.7	1.02	4

[a]Remnant magnetism of the Martian crust appears to require an ancient dynamo with a field strength of at least ten times the Earth's magnetic field.
[b]Since the giant planets do not possess a well-defined boundary of a highly conducting core, values of the most likely region of dynamo activity are given.
Reproduced from Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11; Nellis WJ, Hamilton DC, Holmes NC, Radousky HB, Ree FH, Mitchell AC and Nicol M (1988) The nature of the interior of Uranus based on studies of planetary ices at high dynamic pressure. *Science* 240: 779–781; Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98: 18659–18679; and various websites.

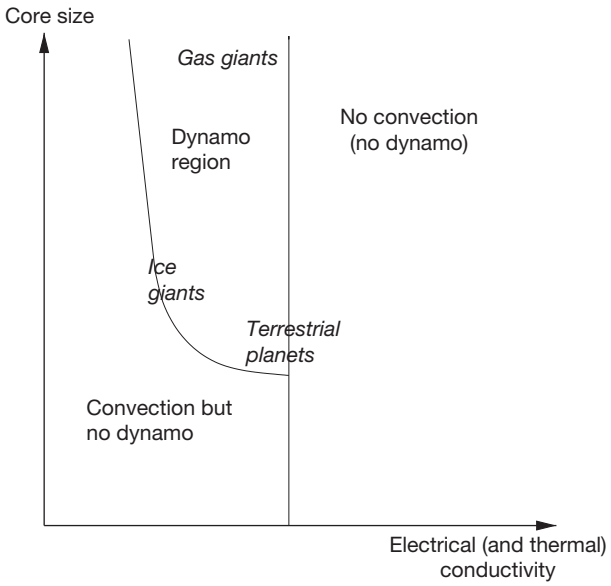


Figure 13 Sketch of Stevenson's paradox. Reproduced from Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.

likely that the dynamos of these planets operate in regions where hydrogen is still a semiconductor. For more details, we refer to the review of Stanley and Glatzmaier (2010).

In the following, brief characterizations of the magnetisms of various planets and moons are given. In terrestrial planets and satellites, dynamos are possible in the liquid parts of their iron cores. Such cores always include light elements that depress the temperature of freezing and can give rise to compositional buoyancy since they stay in solution while a solid inner core is growing. In Jupiter and especially in Saturn, convection may also be driven in part by compositional buoyancy since hydrogen and helium are immiscible in certain regions and helium drops may rain out. In numerical dynamo models, the different sources of buoyancy are usually not distinguished, however.

10.07.6.2 Mercury

It has already become apparent that the origin of Mercury's magnetism is especially enigmatic. Besides the hypothesis of an active dynamo, thermoelectric currents and remnant magnetism of the crust compete as explanations. The latter possibility had been discarded for a long time since Runcorn's theorem states that a homogeneous spherical shell cannot be magnetized by an interior magnetic field in such a way that a dipolar field can be observed from the outside after the interior source has been removed (Runcorn, 1975). Aharonson et al. (2004) had shown, however, that plausible inhomogeneities of the crust could possibly explain a remnant magnetic field created by an ancient dynamo.

The idea that a convection-driven dynamo operates in a fluid outer core of Mercury has received some support from observations of Mercury's librations (Margot et al., 2004). The amplitudes of the small periodic variations (librations) of Mercury's rotation caused by the gravitational pull of the sun on the nonaxisymmetric mass distribution of Mercury appear to be too large for a fully solidified planet. This suggests that the solid inner core is separated from the mantle by a fluid outer core in which convection flows may occur.

According to Williams et al. (2007), a present-day dynamo in Mercury is difficult to achieve because of possibly insufficient heat sources for driving convection. Nevertheless, numerous dynamo models have been proposed, which attempt to explain the rather low amplitude of the Hermian magnetic field. The numerical simulations by Stanley et al. (2005) of dynamos in a thin shell liquid outer core have demonstrated that the dipole strength measured from the outside may not be representative for the strength of the magnetic field at the core–mantle boundary and the criteria mentioned in **Section 10.07.6.1** may be satisfied after all. Christensen (2006) and Christensen and Wicht (2008) advocated a stably stratified outer part of the liquid outer core that effectively damps the field generated by convection in the inner part of the liquid outer core. The model has recently been improved by Manglik et al. (2010) who have taken into account compositional and thermal buoyancy separately. When the compositional buoyancy plays a minor role, it can be demonstrated that Mercury's dynamo may fit the scaling relationship [20] of Christensen and Aubert (2006).

Heyner et al. (2011) showed through numerical simulations that a negative feedback between the magnetospheric and the internal magnetic fields can exist, which would dampen the latter. The question, however, whether there is a sufficiently strong source of buoyancy to sustain convection and the dynamo over the age of the planet has not yet been answered satisfactorily.

The recent measurements of the MESSENGER spacecraft have permitted a much improved determination of Mercury's magnetic field. It resembles a nearly axisymmetric dipole field but with a dipole origin displaced northward from the center of the planet such that the field strength at the North Pole exceeds that at the South Pole by a factor of 3.4 (Anderson et al., 2011). Such a configuration may remind one of a hemispheric dynamo as discussed in the context of **Figures 8** and 9.

10.07.6.3 Venus

Space probes have clearly shown that presently, there is no dynamo operating in Venus. This result is surprising since Venus is very similar to the Earth in many other aspects. That the period of rotation is only 1/243 of a day should not matter much since the Coriolis parameter τ is still huge. More important is the fact that the cooling of the planet is less efficient because of the apparent absence of mantle convection and plate tectonics. As model calculations by Stevenson et al. (1983) have shown, the existence of a dynamo in the core of a terrestrial planet depends sensitively on the thermal history of the planet. In Venus, a solid inner core may not yet have started to grow, and sufficiently vigorous convection does not seem to occur in the core, at least not at the present time. Venus could have had a dynamo in the past, perhaps until 1.5 Gy ago, and a future dynamo may well be possible.

10.07.6.4 Mars

Mars does not have an active dynamo, but the strong magnetization of parts of the Martian crust in its Southern Hemisphere requires a field of at least ten times the strength of the present geomagnetic field. The observed north–south dichotomy of the Martian crustal magnetization has given rise to rather different theoretical explanations. According to Milbury and Schubert (2010), large parts of the originally fairly homogeneously magnetized crust have later been demagnetized or removed through impacts. On the other hand, the asymmetry of the magnetization could be a consequence of an asymmetry of the convection flows in the Martian iron core (Stanley et al., 2008). Amit et al. (2011) demonstrated with their dynamo models that a moderate heat flux anomaly caused by a mantle heterogeneity at the geographic pole may be sufficient to produce the magnetization dichotomy. There is general agreement that the Martian dynamo operated before 4 Gy ago (Lillis et al., 2008). It may have been stopped by a giant impact heating the Martian core through shock waves (Arkani-Hamed and Olson, 2010). Other authors argue that the Martian dynamo operated in the Noachian period (about 4.1–3.7 Gy ago) and perhaps even beyond that time. See discussion by Schubert and Soderlund (2011).

10.07.6.5 Moon

Like Mars, the Moon does not have an active dynamo, but the magnetization of the lunar crust detected by the Apollo subsatellites and the paleomagnetic analysis of lunar rocks indicate that a dynamo has been operating, at least intermittently, in the time span between 4.2 and 3.7 Gy ago (Shea et al., 2012). The rather small size of the lunar iron core with a radius of about 350 km suggests that a convection-driven dynamo operating throughout a period of 500 My is rather unlikely. An alternative possibility is an intermittently operating dynamo driven by motions in the liquid lunar core generated by major impacts (Dwyer et al., 2011; Le Bars et al., 2011).

10.07.6.6 Jupiter

The Jovian magnetic field is rather similar to the geomagnetic one sharing with it a dipole axis that is inclined by an angle of the order of 10° with respect to the axis of rotation. That the higher harmonics of the field are relatively stronger than those in the case of the Earth suggests that the Jovian dynamo is driven at a more shallow layer than the Earth's outer core. This assumption is in accordance with the transition to metallic hydrogen, which is expected to occur at a pressure of 140 GPa corresponding to a depth of about 15% of the Jovian radius (Nellis et al., 1996). The transition is not a sharp one as suggested by earlier models of Jupiter's interior, but a gradual one leading to a decrease of the magnetic diffusivity down to about 4 $m^2\,s^{-1}$. Since this value yields a rather high value for the Elsasser number Λ, Stevenson (2003) had argued that the dynamo is most active above the region of highest conductivity. Jones (2003), on the other hand, accepted the value 20 for Λ but must admit a high magnetic Reynolds number of the order 10^4 at which an effective dynamo may no longer be possible.

A detailed model of a convection-driven Jovian dynamo should include effects of compressibility and depth-dependent electrical conductivity. A first attempt in this direction has most recently been published by Duarte et al. (2013). In this paper, the somewhat controversial ideas about the penetration of the surface differential rotation into the Jovian interior and the location of dynamo action are also discussed.

10.07.6.7 Saturn

Saturn's magnetic field with a strength of 0.2 mT near the poles is much weaker than that of Jupiter, which must be attributed primarily to its deeper origin in the planet. The transition pressure of 140 GPa is reached only at a depth of about 40% of the planetary radius. A property that has received much attention is that the Saturnian magnetic field is nearly axisymmetric with respect to the axis of rotation. This property cannot be interpreted as a contradiction to Cowling's (1934) theorem since it has long been shown that fields with arbitrarily small deviations from axisymmetry can be generated by the dynamo process (Braginsky, 1976). Stevenson (1982) gave an explanation for the almost axisymmetric Saturnian field by demonstrating that the differential rotation in the stably stratified but still electrically conducting gas shells above the dynamo region will tend to shear off all nonaxisymmetric components of the field. The problem may be a bit more intricate, as has been pointed out by Love (2000) who showed that sometimes, fields with solely nonaxisymmetric components are found in such situations.

10.07.6.8 Uranus and Neptune

A metallic liquid probably does not exist in Uranus and Neptune whose interiors consist mostly of ice and rocks. Because of the pressure dissociation of water, it is believed that deeper regions of these planets are characterized by an ionic conductivity that is lower than typical metallic conductivities by between one and two orders of magnitude. Recent laboratory measurements (Lee et al., 2006) suggest, however, that water is semiconducting at pressures of 100 GPa and relevant temperatures of the order of 4000 K. The mixture of water, methane, ammonia, and other ices is sufficiently fluid that convection can occur and that a dynamo is thus possible. The fact that the observed magnetic fields do not show an alignment with the axis of planetary rotation and that quadrupolar and octupolar components are comparable to the dipolar components of the fields has been attributed to a dynamo operating in a thin shell (Ruzmaikin and Starchenko, 1991; Stanley and Bloxham, 2004). This seems to be hardly necessary, however, since convection-driven dynamos in thick shells often exhibit quadrupolar or hemispheric structures (Grote and Busse, 2001; Grote et al., 2000). In particular, such fields are found when small-scale components dominate as in the case of high Rayleigh numbers and low Prandtl numbers. See, for example, **Figure 11** that shows a magnetic field rather similar to that of Uranus as displayed in the paper of Connerney (1993). A special difficulty for a convection-driven dynamo in the interior of Uranus is caused by its low emission of heat. Holme and Bloxham (1996) suggested that a typical dynamo would involve more Ohmic dissipation than that corresponds to the heat flux emitted from the interior of the planet.

10.07.6.9 Ganymede and Other Satellites

It came as a great surprise when the measurements by the Galileo spacecraft indicated that Jupiter's moon Ganymede possesses a global magnetic field for which an active dynamo inside the satellite seems to be the only realistic explanation (Kivelson et al., 1996). The only competing explanation that cannot easily be excluded is remnant magnetism (Crary and Bagenal, 1998). It must be kept in mind that Ganymede is the largest satellite in the solar system, which exceeds even Mercury in size. Nevertheless, the estimated radius of its iron core is only about 660 km, and it is hard to believe that sufficiently strong convection could still exist in the core unless it contains a lot of radioactive potassium 40. Even a capture of Ganymede into a resonance in its more recent history would not have supplied sufficient heating (Showman and Malhotra, 1999). But see also the discussion by Breuer et al. (2010).

The fact that Ganymede's magnetic moment is nearly aligned with the ambient Jovian magnetic field has led to the suggestion that this field could aid significantly Ganymede's dynamo. But Sarson et al. (1997) had investigated this question with the result that the ambient field is too weak to exert much influence.

In the case of Io that is much closer to Jupiter, the ambient Jovian field could eventually aid a dynamo, but there is no need for such an assumption since the observations obtained by the Galileo spacecraft can readily be explained by the interaction of the Jovian field with the hot electrically conducting interior of Io (Khurana et al., 2011).

There are no other satellites in the solar system where an active dynamo must be suspected. The substantial magnetization of many meteorites suggests, however, that several differentiated protoplanets have had dynamos in the early days of the solar system. This fact is also supported by the remanent magnetism that has been measured at several asteroids.

10.07.7 Concluding Remarks

It is apparent from the discussions in the preceding text that dynamo theory does not yet have much specific information to contribute to the interpretation of the observed magnetic properties of planets and satellites in the solar system. Even possible interactions between thermal and compositional buoyancies have not yet been taken fully into account (Breuer et al., 2010; Busse, 2002b; Glatzmaier and Roberts, 1996; Simitev, 2011). Some typical properties are already apparent, however, such as the following:

- Dynamos that are dominated by a nearly axial dipole and exhibit a magnetic energy that exceeds the kinetic one by orders of magnitude as in the case of the Earth are typical for highly effective Prandtl numbers as must be expected when convection is primarily driven by compositional buoyancy. The rather Earth-like appearance of the dynamo of Glatzmaier and Roberts (1995) is in part due to the high value of P used in their computational model.
- Dynamos exhibiting strong higher harmonics are more likely driven by thermal convection corresponding to the Prandtl numbers of the order unity or less.
- Dynamo oscillations are a likely phenomenon in the presence of a sufficiently strong differential rotation. They may not always be apparent in the poloidal magnetic field observed at a distance from the dynamo region.
- Considerations based on the effects on convection of an imposed homogeneous magnetic field cannot directly be applied to the case of convection-driven dynamos.

More details on the parameter dependence of dynamos will certainly emerge in the future as the increasingly available computer capacity will allow extensions of the parameter space accessible to computer simulations. Future space probes will provide much needed detailed information on planetary magnetic fields. Eventually, we may learn more about their variation in time, which is one of the most interesting properties of planetary magnetic fields.

As in the case of stellar magnetism where the study of star spots and stellar magnetic cycles is contributing to the understanding of solar magnetism, it may eventually become possible to learn about the magnetism of extrasolar planets and apply this knowledge for an improved understanding of solar system dynamos. Undoubtedly, the field of planetary magnetism will continue to be an exciting one!

References

Aharonson O, Zuber MT, and Solomon SC (2004) Crustal remanence in an internally magnetized non-uniform shell: A possible source for Mercury's magnetic field? *Earth and Planetary Science Letters* 218: 261–268.

Amit H, Christensen UR, and Langlais B (2011) The influence of degree-1 mantle heterogeneity on the past dynamo of Mars. *Physics of the Earth and Planetary Interiors* 189: 63–79.

Anderson BJ, Johnson CL, Korth H, et al. (2011) The global magnetic field of Mercury from MESSENGER orbital observations. *Science* 333: 1859–1862.

Ardes M, Busse FH, and Wicht J (1997) Thermal convection in rotating spherical shells. *Physics of the Earth and Planetary Interiors* 99: 55–67.

Arkani-Hamed J and Olson P (2010) Giant impact stratification of the Martian core. *Geophysical Research Letters* 37, L02201.

Aubert J, Labrosse S, and Poitou C (2009) Modelling the palaeo-evolution of the geodynamo. *Geophysical Journal International* 179: 1414–1428.

Backus G (1958) A class of self-sustaining dissipative spherical dynamos. *Annals of Physics* 4: 372–447.

Braginsky SI (1967) Magnetic waves in the Earth's core. *Geomagnetism and Aeronomy* 7: 851–859.

Braginsky SI (1976) On the nearly axially-symmetrical model of the hydromagnetic dynamo of the Earth. *Physics of the Earth and Planetary Interiors* 11: 191–199.

Breuer M, Manglik A, Wicht J, Truemper T, Harder H, and Hansen U (2010) Thermochemically driven convection in a rotating spherical shell. *Geophysical Journal International* 183: 150–162.

Brun AS, Miesch MS, and Toomre J (2004) Global-scale turbulent convection and magnetic dynamo action in the solar envelope. *Astrophysics Journal* 614: 1073–1098.

Bullard EC (1949) The magnetic field within the Earth. *Proceedings of the Royal Society of London A* 197: 433–463.

Burke BF and Franklin KL (1955) Observations of variable radio source associated with the planet Jupiter. *Journal of Geophysical Research* 60: 213–217.

Busse FH (1970) Differential rotation in stellar convection zones. *Astrophysics Journal* 159: 629–639.

Busse FH (1976) Generation of planetary magnetism by convection. *Physics of the Earth and Planetary Interiors* 12: 350–358.

Busse FH (1983) Recent developments in the dynamo theory of planetary magnetism. *Annual Review of Earth and Planetary Sciences* 11: 241–268.

Busse FH (2002a) Convection flows in rapidly rotating spheres and their dynamo action. *Physics of Fluids* 14: 1301–1314.

Busse FH (2002b) Is low Rayleigh number convection possible in the Earth's core? *Geophysical Research Letters* 29, GL0149597.

Busse FH and Carrigan CR (1976) Laboratory simulation of thermal convection in rotating planets and stars. *Science* 191: 81–83.

Busse FH and Cuong PG (1977) Convection in rapidly rotating spherical fluid shells. *Geophysical and Astrophysical Fluid Dynamics* 8: 17–41.

Busse FH, Grote E, and Simitev R (2003) Convection in rotating spherical shells and its dynamo action. In: Jones CA, Soward AM, and Zhang K (eds.) *Earth's Core and Lower Mantle*, pp. 130–152. London, UK: Taylor & Francis.

Busse FH and Heikes KE (1980) Convection in a rotating layer: A simple case of turbulence. *Science* 208: 173–175.

Busse FH and Simitev R (2004) Inertial convection in rotating fluid spheres. *Journal of Fluid Mechanics* 498: 23–30.

Busse FH and Simitev R (2008) Toroidal flux oscillation as possible cause of geomagnetic excursions and reversals. *Physics of the Earth and Planetary Interiors* 168: 237–243.

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon Press.

Childress S and Soward AM (1972) Convection-driven hydrodynamic dynamo. *Physical Review Letters* 29: 837–839.

Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.

Christensen UR and Aubert J (2006) Scaling laws for dynamos in rotating spherical shells. *Geophysical Journal International* 166: 97–114.

Christensen UR, Aubert J, Cardin P, et al. (2001) A numerical dynamo benchmark. *Physics of the Earth and Planetary Interiors* 128: 25–34.

Christensen U, Olson P, and Glatzmaier GA (1999) Numerical modeling of the geodynamo: A systematic parameter study. *Geophysical Journal International* 138: 393–409.

Christensen UR and Wicht J (2008) Models of magnetic field generation in partly stable planetary cores: Applications to Mercury and Saturn. *Icarus* 196: 16–34.

Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98: 18659–18679.

Cowling TG (1934) The magnetic field of sunspots. *Monthly Notices of the Royal Astronomical Society* 34: 39–48.

Crary FJ and Bagenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research* 103: 25757–25773.

Dolginov Sh (1977) Planetary magnetism: A survey. *Geomagnetism and Aeronomy* 17: 391–406.

Duarte L, Gastine T, and Wicht J (2013) Anelastic dynamo models with variable electrical conductivity: An application to gas giants. *Physics of the Earth and Planetary Interiors* 222: 22–34.

Dwyer CA, Stevenson DJ, and Nimmo F (2011) A long-lived lunar dynamo driven by continuous mechanical stirring. *Nature* 479: 212–214.

Fearn DR (1979) Thermal and magnetic instabilities in a rapidly rotating fluid sphere. *Geophysical and Astrophysical Fluid Dynamics* 14: 103–126.

Giampieri G and Balogh A (2002) Mercury's thermoelectric dynamo model revisited. *Planetary and Space Science* 50: 757–762.

Gilbert N (1600) *De Magnete, Gilbert Club revised English translation*. London: Chiswick Press.

Glatzmaier GA and Roberts PH (1995) A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* 377: 203–209.

Glatzmaier GA and Roberts PH (1996) An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection. *Physica D* 97: 81–94.

Grote E and Busse FH (2001) Dynamics of convection and dynamos in rotating spherical fluid shells. *Fluid Dynamics Research* 28: 349–368.

Grote E, Busse FH, and Tilgner A (2000) Regular and chaotic spherical dynamos. *Physics of the Earth and Planetary Interiors* 117: 259–272.

Herzenberg A (1958) Geomagnetic dynamos. *Philosophical Transactions of the Royal Society of London* A250: 543–585.

Heyner D, Wicht J, Gomez-Perez N, Schmitt D, Auster H-U, and Glassmeier K-H (2011) Evidence from numerical experiments for a feedback dynamo generating Mercury's magnetic field. *Science* 334: 1690–1693.

Hide R (1974) Jupiter and Saturn. *Proceedings of the Royal Society of London A* 336: 63–84.

Holme R and Bloxham J (1996) The magnetic fields of Uranus and Neptune: Methods and models. *Journal of Geophysical Research* 101: 2177–2200.

Jacobs JA (1979) Planetary magnetic fields. *Geophysical Research Letters* 6: 213–214.

Jones CA (2000) Convection-driven geodynamo models. *Philosophical Transactions of the Royal Society of London* A358: 873–897.

Jones CA (2003) Dynamos in planets. In: Thompson MJ and Christensen-Daalsgard J (eds.) *Stellar Astrophysical Fluid Dynamics*, pp. 159–176. Cambridge: Cambridge University Press.

Jones CA (2011) Planetary magnetic fields and fluid dynamos. *Annual Review of Fluid Mechanics* 43: 583–614.

Jones CA, Boronski P, Brun AS, et al. (2011) Anelastic convection-driven dynamo benchmarks. *Icarus* 216: 120–135.

Jones CA, Soward AM, and Mussa AI (2000) The onset of thermal convection in a rapidly rotating sphere. *Journal of Fluid Mechanics* 405: 157–179.

Khurana KK, Jia X, Kivelson MG, Nimmo F, Schubert G, and Russell CT (2011) Evidence of a global magma ocean in Io's interior. *Science* 332: 1186–1189.

Kivelson MG, Khurana KK, Russell CT, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.

Kutzner C and Christensen UR (2000) Effects of driving mechanisms in geodynamo models. *Geophysical Research Letters* 27: 29–32.

Kutzner C and Christensen UR (2002) From stable dipolar towards reversing numerical dynamos. *Physics of the Earth and Planetary Interiors* 131: 29–45.

Larmor J (1919) How could a rotating body such as the sun become a magnet? *British Association for the Advancement of Science Report*: 159–160.

Le Bars M, Wieczorek MA, Karatekin O, Cebron D, and Laneuville M (2011) An impact-driven dynamo for the early Moon. *Nature* 479: 215–218.

Lee KKM, Benedetti LR, Jeanloz R, et al. (2006) Laser-driven shock experiments on precompressed water: Implications for "icy" giant planets. *Journal of Chemical Physics* 125: 014701.

Lillis RJ, Frey HV, and Manga M (2008) Rapid decrease in martian crustal magnetization in the Noachian era: Implications for the dynamo and climate of early Mars. *Geophysical Research Letters* 35, L14203.

Liu J, Goldreich PM, and Stevenson DJ (2008) Ohmic dissipation constraints on deep-seated zonal winds in Jupiter and Saturn. *Icarus* 196: 597–615.

Love JJ (2000) Dynamo action and the nearly axisymmetric magnetic field of Saturn. *Geophysical Research Letters* 27: 2889–2892.

Malkus WVR (1968) Precession of the Earth as the cause of geomagnetism. *Science* 160: 259–264.

Malkus WVR (1994) Energy sources for planetary dynamos. In: Proctor MRE and Gilbert AD (eds.) *Lectures on Planetary and Solar Dynamos*, pp. 161–179. Cambridge: Cambridge University Press.

Manglik A, Wicht J, and Christensen UR (2010) A dynamo model with double diffusive convection for Mercury's core. *Earth and Planetary Science Letters* 289: 619–628.

Margot J, Peale S, Jurgens RF, Slade MA, and Holin IV (2004) Earth-based measurements of Mercury's forced librations in longitude. *Eos, Transactions American Geophysical Union* 85, GP33–03.

Milbury C and Schubert G (2010) Search for the global signature of the Martian dynamo. *Journal of Geophysical Research* 115: E10010.

Nellis WJ, Weir ST, and Mitchell AC (1996) Metallization and electrical conductivity of hydrogen in Jupiter. *Science* 273: 936–938.

Runcorn S (1975) An ancient lunar magnetic field. *Nature* 253: 701–703.

Ruzmaikin AA and Starchenko SV (1991) On the origin of Uranus and Neptune magnetic fields. *Icarus* 93: 82–87.

Sarson GR, Jones CA, Zhang K, and Schubert G (1997) Magnetoconvection dynamos and the magnetic field of Io and Ganymede. *Science* 276: 1106–1108.

Schubert G and Soderlund KM (2011) Planetary magnetic fields: Observations and models. *Physics of the Earth and Planetary Interiors* 187: 92–108.

Shea EK, Weiss BP, Cassata WS, et al. (2012) A long-lived lunar dynamo. *Science* 335: 453–456.

Showman AP and Malhotra R (1999) The Galilean satellites. *Science* 286: 77–84.

Simitev R (2011) Doubly-diffusive convection in a rotating cylindrical annulus with conical caps. *Physics of the Earth and Planetary Interiors* 186: 183–190.

Simitev R and Busse FH (2005) Prandtl-number dependence of convection-driven dynamos in rotating spherical fluid shells. *Journal of Fluid Mechanics* 532: 365–388.

Simitev R and Busse FH (2009) Bistability and hysteresis of dipolar dynamos generated by turbulent convection in rotating spherical shells. *EPL* 85: 19001.

Simitev R and Busse FH (2012) Bistable attractors in a model of convection-driven spherical dynamos. *Physica Scripta* 86: 018409.

Soderlund KM, King EM, and Aurnou JM (2012) The influence of magnetic fields in planetary dynamo models. *Earth and Planetary Science Letters* 333–334: 9–20.

Somerville RCJ and Lipps FB (1973) A numerical study in three space dimensions of Bénard convection in a rotating fluid. *Journal of the Atmospheric Sciences* 30: 590–596.

Stanley S and Bloxham J (2004) Convective-region geometry as the cause of Uranus' and Neptune's magnetic fields. *Nature* 428: 151–153.

Stanley S, Bloxham J, Hutchison WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.

Stanley S, Elkins-Tanton L, Zuber MT, and Parmentier EM (2008) Mars' paleomagnetic field as the result of a single-hemisphere dynamo. *Science* 321: 1822–1825.

Stanley S and Glatzmaier G (2010) Dynamo models for planets other than earth. *Space Science Reviews* 152: 617–649.

Stevens RJAM, Overkamp J, Lohse D, and Clercx JH (2011) Effect of aspect ratio on vortex distribution and heat transfer in rotating Rayleigh–Bénard convection. *Physical Review E* 84: 056313.

Stevenson DJ (1979) Turbulent thermal convection in the presence of rotation and a magnetic field. *Geophysical and Astrophysical Fluid Dynamics* 12: 139–169.

Stevenson DJ (1982) Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophysical and Astrophysical Fluid Dynamics* 21: 113–127.

Stevenson DJ (1983) Planetary magnetic fields. *Reports on Progress in Physics* 46: 555–620.

Stevenson DJ (1984) The energy flux number and three types of planetary dynamo. *Astronomische Nachrichten* 305: 257–264.

Stevenson DJ (1987) Mercury's magnetic field: A thermoelectric dynamo? *Earth and Planetary Science Letters* 82: 114–120.

Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.

Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.

Tilgner A (1999) Spectral methods for the simulation of incompressible flows in spherical shells. *International Journal for Numerical Methods in Fluids* 30: 713–724.

Tilgner A (2007) Kinematic dynamos with precession driven flow in a sphere. *Geophysical and Astrophysical Fluid Dynamics* 101: 1–9.

Tilgner A and Busse FH (1997) Finite amplitude convection in rotating spherical fluid shells. *Journal of Fluid Mechanics* 332: 359–376.

Vanyo JP (1984) Earth core motions: Experiments with spheroids. *Geophysical Journal of the Royal Astronomical Society* 77: 173–183.

Williams J-P, Aharonson O, and Nimmo F (2007) Powering Mercury's dynamo. *Geophysical Research Letters* 34, L21201.

Zhang K (1994) On coupling between the Poincaré equation and the heat equation. *Journal of Fluid Mechanics* 268: 211–229.

Zhang K (1995) On coupling between the Poincaré equation and the heat equation: No-slip boundary condition. *Journal of Fluid Mechanics* 284: 239–256.

10.08 Dynamics and Thermal History of the Terrestrial Planets, the Moon, and Io

D Breuer, DLR, Institut für Planetenforschung, Berlin, Germany
WB Moore, Hampton University, Hampton, VA, USA

© 2015 Elsevier B.V. All rights reserved.

10.08.1	**Introduction**	255
10.08.2	**Physical and Chemical Properties of Planets and Planetary Materials Bearing on Mantle Dynamics and Thermal Evolution Models**	256
10.08.2.1	Interior Structure and Phase Transitions	256
10.08.2.2	Heat Sources	257
10.08.2.2.1	Accretional heating	257
10.08.2.2.2	Core formation	258
10.08.2.2.3	Radioactive heat sources	258
10.08.2.2.4	Tidal dissipation	259
10.08.2.3	Rheology	259
10.08.2.4	Thermal Expansivity and Conductivity	259
10.08.3	**Planform of Convection**	260
10.08.4	**Thermal Evolution Models Using Parameterized Convection**	262
10.08.4.1	Parameterizations Based on Constant Viscosity Laws	262
10.08.4.2	Stagnant Lid Parameterization Based on Temperature-Dependent Viscosity Laws	263
10.08.4.3	Plate Tectonics, Lithosphere Delamination, and Stagnant Lid Convection	264
10.08.5	**Thermal Evolution, Volcanic History, and Magnetic Field History of Terrestrial Planets**	266
10.08.5.1	Crustal Formation	267
10.08.5.1.1	Primary and secondary crustal formation	267
10.08.5.1.2	Plume volcanism and partial melt zone	268
10.08.5.1.3	Thermal properties of the crust	269
10.08.5.1.4	Observations	269
10.08.5.1.5	Incorporation of crustal growth in thermochemical evolution models	270
10.08.5.2	Magnetic Field Generation	271
10.08.5.2.1	Core convection and necessary conditions for dynamos	271
10.08.6	**Comparison of the Terrestrial Planets and the Moon**	272
10.08.6.1	Mercury	273
10.08.6.2	Venus	276
10.08.6.3	The Moon	279
10.08.6.4	Mars	284
10.08.7	**Io**	291
10.08.7.1	Tidal Heating and Thermal Evolution	291
10.08.7.2	Thermal and Orbital Evolution	293
10.08.8	**Summary**	294
Acknowledgments		295
References		295

10.08.1 Introduction

Unraveling a terrestrial planet's evolution in order to understand the observed processes and features is a difficult task that requires the integration of evidence from a wide variety of fields into consistent models. Geology, geophysics, mineralogy, and cosmochemistry, as well as chemistry, physics, and even biology, all influence the thermal and chemical evolution of terrestrial bodies. Our knowledge of the other terrestrial bodies stems not only mostly from remote sensing satellites but also, in some cases, from in situ chemical measurements and even from surface samples, in the case of the Moon, and meteorites – in the case of the Moon and Mars. These samples have been of immense value for the Moon and Mars science as they provide information not only about the composition but also about processes such as the formation of the core and the crust of those bodies. A fundamental process in planetary evolution is the thermal evolution of its interior, which depends on the internal heat sources, on thermophysical properties, and, to a large extent, on the dynamics of the planet's mantle through which heat is transported by thermal convection. It has been understood since the beginning of the twentieth century that thermal convection is capable of driving mantle circulation. If the mantle is heated from within (or from below) and is cooled from above, it can become gravitationally unstable and thermal convection can occur as colder

rock descends into the mantle and hotter rock ascends toward the surface. This circulation of material transports heat toward the planet's surface and tends to cool the interior, while heat produced within, for example, by the decay of radioactive elements, tends to warm it. The motions driven by convective heat transport result in surface stresses and deformation, producing many geologic features observed on the terrestrial bodies today. An understanding of planetary thermal evolution is critical for interpreting these surface features in terms of interior processes including the dynamics of the mantle and the core.

The plate tectonics of the Earth are a unique expression of mantle convection among the terrestrial planets. Our planet's otherwise stiff lithosphere is segmented by weak zones into distinct plates. Where plates diverge, new lithosphere is created, and where they converge, the lithosphere descends into the mantle. This recycling process transfers heat out of the interior very effectively. One consequence of this process is that the Earth has a very young oceanic crust, generally less than 200 million years old (e.g., Condie, 1997). In addition to the oceanic crust, a less dense continental crust exists that usually does not recycle into the interior and has an average age of about 2000 Ma (e.g., Kemp and Hawkesworth, 2003). Comparing the Earth with the other terrestrial planets, the most striking difference at first glance is the state of the surface (including the crust), which is indicative of the way heat is transported. The surfaces of other planets are not segmented but consist of single plates, the so-called stagnant lids, beneath which the mantle convects. Heat flow through the lid is mainly by conduction, with some minor contribution by volcanic heat transport through this stagnant lid. The difference in the heat transport mechanism for the planets is also reflected in their thermal evolution.

In this chapter, we describe the thermal evolution and the mantle dynamics and related processes like crustal formation and the magnetic field evolution of the terrestrial bodies other than the Earth, that is, Mercury, Venus, Mars, and the Moon. The Galilean satellite Io is included in this chapter since this body consists, like the terrestrial planets and the Moon, of an iron-rich core and a silicate outer shell. The close orbit about Jupiter and the resonant orbital interaction among Io, Europa, and Ganymede make Io, however, special among the other terrestrial bodies. In its interior, tidal deformation plays an important role and provides an extremely large internal heat source. The consequence is a very different thermal evolution for Io than for the other bodies; therefore, we discuss Io in a separate section of this chapter.

The chapter is organized as follows: First, we describe the relevant physical and chemical properties of planets and planetary materials bearing on mantle dynamics and thermal evolution models. In the next section, the concept of mantle convection is introduced and the relationships between the convective parameters and the planform of the flow are discussed. A description of parameterized convection models follows with a comparison of three main heat transport mechanisms: plate tectonics, stagnant lid convection, and lithosphere delamination. In the next section, the crustal formation and magnetic field generation of terrestrial planets are described since both strongly depend on the thermal evolution of a planet and can be used as constraints for the models. The chapter concludes with separate reviews of the thermochemical evolution and mantle dynamics of Mercury, Venus, the Moon, Mars, and Io and includes a summary section.

10.08.2 Physical and Chemical Properties of Planets and Planetary Materials Bearing on Mantle Dynamics and Thermal Evolution Models

10.08.2.1 Interior Structure and Phase Transitions

The interior dynamics and thermal evolution of the terrestrial planets depend strongly on internal structure and composition. One important characteristic is the size and density of the major reservoirs, that is, the core, mantle, and crust. Lacking seismic data for most bodies, the basic structure can be estimated from the mean density and the polar moment of inertia (MoI) factor of the terrestrial bodies. The latter parameter can be obtained from the measurements of the low-order gravity field (J_2 and C_{22}), the figure, and/or the precession rate of the planet. The use of shape or gravity data to infer the MoI requires a planet to be in hydrostatic equilibrium. The determination of the interior structure does not provide a unique solution; there is, for example, a trade-off between the core size and core density. However, with additional constraints on composition, it is possible to derive estimates of the structure (see **Chapter 10.02**). For mantle dynamics, the size of the metallic core is of particular interest. A large core (and hence a thin mantle) suggests less radiogenic heat sources (which tend to be found in the rock phase) and thus lower temperatures of the mantle layer, and one finds in general a convection pattern that is smaller in scale compared to a planet with a thick mantle.

Thermal convection can be strongly influenced by chemical layering and phase transitions in the mantle. In the case of chemical layering, density variations that influence convection are caused not only by temperature variations but also by differences in mantle composition. Chemical layering can block or retard mantle flow if the compositional density increases with depth. In this stable configuration, thermal convection is only possible if density changes due to temperature variations can overcome the compositional density differences. Chemical layering can enhance mantle flow if the compositional component of density decreases with depth. Reshuffling of this unstable configuration results in a stable density configuration unless efficient mixing prevents the establishment of distinct layers. It is also possible that thermal convection takes place in separate and chemically distinct layers. The heat transport in a mantle with layered thermal convection is less efficient than in a mantle with whole mantle convection (Spohn and Schubert, 1982). Chemical stratification resulting from early differentiation of the mantle is one way to explain the presence of separate and chemically distinct reservoirs that might be needed to explain geochemical observations. The density contrast between the layers must be sufficient to overcome the tendency of convective motions to mix the material at the boundary (Gurnis, 1986). At present, we do not know whether distinct chemical reservoirs other than the crust, mantle, and core exist in the terrestrial planets and the Moon.

Phase transitions in planetary mantles also play an important role in the dynamics and thermal evolution of the planet. At a certain pressure and temperature, mantle materials undergo isochemical phase changes, associated with a change in density. The variation of the phase transition temperature T with pressure P is related to the volume and entropy changes,

ΔV and ΔS, respectively, through the Clausius–Clapeyron equation ($dP/dT = \Delta S/\Delta V$). The phase change from olivine to spinel has a positive Clapeyron slope (also called exothermic phase transition in the literature). Thus, for downwelling cold material, the depth of the phase change is elevated, while for hot upwelling material, the boundary is deflected downward; both distortions cause an acceleration of the moving material. These distortions and therefore the acceleration of the flowing mantle material at the phase transitions are reduced by the release and consumption of latent heat for the rising and sinking material, respectively.

In the case of a negative Clapeyron slope as for the spinel-to-perovskite transition (also called endothermic phase transition in the literature), the mantle flow is impeded or even arrested because the sinking cold material deflects the boundary downward, resulting in a decrease of the local buoyancy. Latent heat that is adsorbed by the sinking material at the phase transition may partially compensate the boundary deflection and reduce the blocking mechanism. Likewise, upwelling material may be impeded or arrested in a similar manner.

Depending on the pressure gradient within a planet, these transitions are located at different depths. For instance, in the Earth's mantle, the olivine-to-spinel transitions are located at a depth of 400–520 km, whereas in Venus, they occur at 450–580 km, and in the smaller Mars, the depth of this transition zone is roughly between 1000 and 1500 km. Mineral physics experiments and ab initio calculations (Murakami et al., 2004; Oganov and Ono, 2004; Tsuchiya et al., 2004) have indicated that perovskite undergoes an exothermic phase change to a postperovskite structure just above the core–mantle boundary (CMB) of the Earth. For Mercury and the Moon, the pressure gradient is too low, and olivine-to-spinel phase transition pressures are not reached in their mantles. Details of the interior structure of the terrestrial planets and the Moon are given in Chapter 10.02.

10.08.2.2 Heat Sources

The planet's internal heat sources provide the energy for its dynamics, supplying it with the driving energy for mantle convection, and for melting the mantle that may lead to volcanism. The internal heat energy of the terrestrial planets was greater in the early stages of their histories than it is today, having accumulated rapidly by heat conversion associated with three separate processes, all of which were most intense during the first few million years of the planet's history: (1) the accretion of the planet by impacts, (2) core formation and the associated release of gravitational potential energy, and (3) the radioactive decay of unstable isotopes. Another important heat source in a terrestrial body can be tidal dissipation, which may have been important in Earth's early history and will be seen to dominate in the case of Io, later in the text. This heat source is not necessarily higher early in the evolution; rather, it can be highly variable both spatially and temporally and depends on the body's orbital and thermal evolution.

10.08.2.2.1 Accretional heating

Accretion of material during the formation of planets is likely one of the largest sources of heat. The accretional heat is the energy accumulated through the burial of heat by impacts as the planet grows by the accretion of planetesimals. During accretion, a temperature profile is generated for which the temperature increases from the center toward the surface (Figure 1). A possible postaccretional temperature profile assuming homogeneous accretion of the undifferentiated material and ignoring the contribution of short-lived radiogenic elements like ^{26}Al (e.g., Schubert et al., 1986) is given by

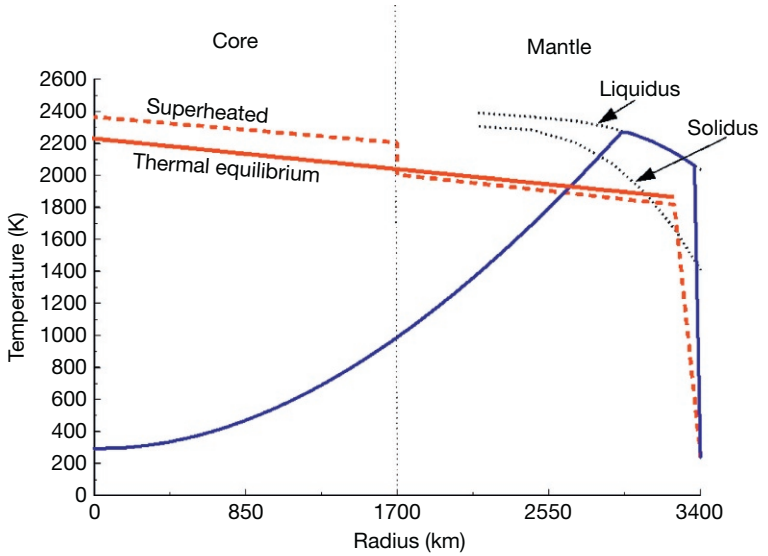


Figure 1 Accretional temperature profile for a planet with the size of Mars (blue solid line) calculated from eqn [1] with a retention factor h of 0.2. The accretional temperature exceeds the solidus and the liquidus close to the surface, thus forming a 600 km deep magma ocean. After core formation, two possible temperature profiles can develop if (1) the core formed in thermal equilibrium with the mantle (solid red line) and (2) the core is superheated with respect to the mantle as the consequence of rapid core formation (dashed red line). The temperature profiles account for the potential energy release due to core formation (eqn [4]). This energy is distributed homogeneously assuming an adiabatic interior.

$$T_a(r) = h\frac{GM(r)}{C_p r}\left(1+\frac{ru^2}{2GM(r)}\right)+T_e+\Delta T_{a,ad} \quad [1]$$

where h is the retention factor describing the contribution of heat retained in the accumulated body, G is the gravitational constant, $M(r)$ is the mass of the planet internal to radius r, C_p is the specific heat, $u^2/2$ is the approach kinetic energy per unit mass, and T_e is the temperature of the surrounding nebula. The postaccretional temperature profile of Schubert et al. (1986) is corrected to account for the temperature rise, $\Delta T_{a,ad}$, caused by adiabatic compression as the internal pressure progressively increases during accumulation.

This accretion temperature distribution is stable against thermal convection and shows an inverse profile, which decreases with depth (**Figure 1**), when compared to a typical temperature profile for planetary interiors, which increases with depth. Depending on the amount of stored heat (part of the energy is radiated back to space) and, in particular, on the size of the body, the accretion profile can rise above the melting temperature of iron and possibly of silicate close to the surface; a magma ocean may form. The temperature distribution due to accretion can also be influenced by predifferentiated planetesimals forming the planet or – if the body formed rapidly enough as suggested for Mars (Dauphas and Pourmand, 2011) – by short-lived radioactive elements such as ^{26}Al. In that case, higher temperatures in the deep interior can be expected.

10.08.2.2.2 Core formation

Planets are assumed to have initially accreted as a mixture of silicate and metal particles. It is, however, likely that some if not all of the planetesimals were already differentiated (e.g., Baker et al., 2005; Kleine et al., 2002) particularly as accretion progresses. The separation of the metal particles, that is, the formation of an iron-rich core, most likely was contemporaneous with or shortly after a planet's accretion. Recent studies on short-lived radionuclides (e.g., ^{182}Hf) suggest that the separation between silicates and iron occurred on the order of $1–5\times10^7$ years after accretion (e.g., Kleine et al., 2002). Estimates of the timescale of the core formation process suggest that for rapid separation, both the silicate and the iron need to be fluid at least in the upper part of the planet (e.g., Stevenson, 1990). In such a scenario, iron accumulates at the base of a magma ocean and sinks further to the center of the planet by Rayleigh–Taylor instability due to iron's larger density. Alternatively, it has been assumed that no global magma ocean formed during accretion; rather, silicate–metal separation takes place without global-scale melting but in magma ponds generated by impacts (Tonks and Melosh, 1992). Numerical studies of this process, however, suggest that the time to form a core would be about 0.5–1 Ga (Senshu et al., 2002) although more recent studies in two- and three-dimensional (2-D and 3-D) show that diapir sinking and thus core formation can be a very fast process (Golabek et al., 2009; Samuel and Tackley, 2008).

The gravitational energy released by core formation is converted into thermal energy through viscous dissipation, which strongly heats the interior. During core formation, when iron sinks to the center, the temperature profile is inverted to decrease from the center toward the surface. The total energy released by the differentiation of a homogeneous planet into an iron-rich core and a silicate mantle can be estimated from the difference between the potential energy stored in a homogeneous planet after accretion, U_H, and the potential energy of the differentiated, two-layer planet, U_D (Schubert et al., 1986). Thus, the energy dissipated by differentiation is

$$U_{Diff} = U_H - U_D \quad [2]$$

with the potential energy stored in a homogeneous body after accretion

$$U_H = -\frac{16}{15}\pi^2 G R_p^5 \rho^2 \quad [3]$$

and with the potential energy of a two-layer planet (with constant density ρ_m in the mantle and ρ_c in the core) given by

$$U_D = -\frac{16}{15}\pi^2 G R^5\left[\rho_m^2+\frac{5}{2}\rho_m(\rho-\rho_m)+\left(\frac{3}{2}\rho_m-\rho_c\right)(\rho_m-\rho_c)\left(\frac{\rho-\rho_m}{\rho_c-\rho_m}\right)^{5/3}\right] \quad [4]$$

If this energy is homogeneously distributed in the planetary interior, that is, if there is thermal equilibrium between the mantle and the core, estimates for the mean temperature increase are about 1000 K for the Earth and about 300 K for Mars (Solomon, 1979). It is, however, likely that due to the assumed rapid core formation process, the core was superheated with respect to the mantle (**Figure 1**). If the core was superheated or not depends strongly on the rheology of the planet and on the segregation mechanisms (Ke and Solomatov, 2009; Samuel and Tackley, 2008).

10.08.2.2.3 Radioactive heat sources

The energy release by accretion and core formation occurred very early in the planet's history since one expects these processes to be very rapid (e.g., Kleine et al., 2002). Thus, they are not responsible for magmatism or tectonics after the first few hundred million years. Instead, long-lived radioactive elements with half-lives of billions of years, primarily ^{238}U, ^{235}U, ^{232}Th, and ^{40}K, become the dominant heat sources. It is this heat that can produce volcanism and power internal convection and tectonics on planets. At present, roughly half of the Earth's heat output is thought to be due to radioactive cooling, while the remainder may be due to secular cooling (loss of sensible heat). For the other terrestrial planets, this ratio (known as the Urey ratio) might be similar. The distribution of the radioactive heat source can change with time as a consequence of chemical differentiation (crustal formation). The radiogenic elements are incompatible in the solid mantle undergoing melting and thus become enriched in the melt. This melt then rises to the surface and forms new crust. If the radioactively enriched crust is stably stratified at the surface, the mantle heat sources are reduced. Short-lived isotopes, in particular ^{26}Al with a half-life of 0.74 Ma, might also have played an important role in heating rocky bodies when the solar system was first forming (e.g., Bizzarro et al., 2004; Lee et al., 1976; MacPherson et al., 1995). The heating rate was high at an early stage but disappeared quickly. Possibly, this strong energy peak was responsible for the differentiation of some of the asteroids (e.g., Ghosh and

McSween, 1998; Merk et al., 2002; Neumann et al., 2012), of planetesimals that formed the planets and early-formed planets (or surviving planetary embryos) such as Mars (Dauphas and Pourmand, 2011).

10.08.2.2.4 Tidal dissipation

An important heat source in terrestrial planets can be tidal energy caused by gravitational interactions between the Sun and the planets and/or between the planets and their satellites. For a satellite that moves on an elliptical orbit, the distance and therefore the gravitational attraction vary periodically. If the interior reacts to this periodic deformation viscoelastically, tidal energy can be dissipated in the form of heat. The heating due to tidal dissipation can, in some cases, exceed the heat released by the decay of radioactive elements. The tidal energy dissipation rate $\dot{E}$ depends on the time variation of the tides, the interior structure, and the rheology (and, hence, temperature) of the dissipating body. Since this energy is derived from the orbital and rotational energies of the interacting bodies, orbital evolution and rotational evolution are therefore coupled to the thermal evolution by the tidal dissipation rate. In the case of a small inclination of the equatorial plane to the orbital plane and a small eccentricity, the tidal dissipation rate in a synchronous satellite can be estimated using the relation by Segatz et al. (1988):

$$\dot{E} \sim \mathrm{Im}(k_2)_{\mathrm{Sat}} \frac{(nR_{\mathrm{Sat}})^5}{G} e^2 \qquad [5]$$

with R_{Sat} the radius of the satellite, n the mean motion, e the eccentricity, $\mathrm{Im}(k_2)_{\mathrm{Sat}}$ the imaginary part of the tidal Love number k_2 (measuring the phase lag of the tide), and G the gravitational constant.

10.08.2.3 Rheology

Rheology is a primary factor affecting convection in terrestrial planets and a rheological prerequisite for convection is that the material behaves on long timescales like a fluid, that is, able to flow under small differential stresses. For mantle rocks, the timescale of deformation is of utmost importance. At high frequency (e.g., the seismic wave band), the mantle is approximately elastic; but at very low frequency (periods larger than 1000 years), the mantle, at sufficient high temperatures and pressure, behaves like a viscous fluid. Unfortunately, such conditions cannot be reproduced in the laboratory and thus the true rheology of planetary mantles is poorly known. What is known is that the viscosity of mantle rocks under the wide variety of conditions in planetary mantles is dependent on temperature, pressure, stress, grain size, and composition (e.g., Karato, 2013).

The exponential dependence of the rheology on the inverse absolute temperature is particularly important in understanding the role of mantle convection in transporting heat. The temperature dependence of the rheology acts as a thermostat to regulate the mantle temperature. Any tendency of the mean temperature to increase is offset by an associated reduction in mantle viscosity, an increase in convective vigor, and a more efficient outward transport of heat. Similarly, a decrease in mantle temperature tends to increase mantle viscosity, reduce convective flow velocities, and decrease the rate of heat transfer. As a result of the sensitive feedback between mantle temperature and rheology, relatively small changes in temperature can produce large changes in heat flux, and the temperature is consequently buffered at the equilibrium point (Tozer, 1965).

The viscosity of the mantle can be written in a simple Arrhenius form as

$$\eta = \eta_0 \exp\left(\frac{A + pV}{RT_{\mathrm{m}}}\right) \qquad [6]$$

In eqn [6], η_0 contains the stress, composition, and grain size-sensitive terms; T_{m} is the mean temperature of the convecting fluid; R is the universal gas constant; A is the activation energy for creep (e.g., Weertman and Weertman, 1975); p is the pressure; and V is the activation volume.

Though planetary mantles are composed of several minerals, olivine is thought to dominate the rheology of silicate mantle rock (e.g., Mackwell, 1991). For dry olivine, the viscosity is about 10^{21} Pa s at a reference temperature of 1600 K (e.g., Karato and Wu, 1993) with an activation energy of about 540 kJ mol^{-1} at pressures of 12 GPa (Karato and Wu, 1993) and an activation volume of about 15 $\mathrm{cm}^3\,\mathrm{mol}^{-1}$ (Karato and Rubie, 1997). This reference viscosity can, for instance, vary with the amount of iron and is a factor of ten lower for the Martian mantle due to its higher iron content in comparison to the Earth's mantle (FeO ~ 17 wt% for Mars vs. FeO ~8 wt% for Mars) (Zhao et al., 2009).

The viscosity is also known to strongly depend on the presence of volatiles. With increasing water content, the viscosity of olivine is reduced by two to three orders of magnitudes in comparison to dry olivine (e.g., Hirth and Kohlstedt, 1996; Mei and Kohlstedt, 2000), resulting in stronger convection for a wet mantle in comparison to a dry mantle at the same temperature and pressure conditions.

10.08.2.4 Thermal Expansivity and Conductivity

The two important parameters for mantle dynamics and thermal evolution of terrestrial planets are the thermal expansivity and the thermal conductivity. The former parameter describes the volume increase per unit temperature increase, that is, the decrease of density with increase of temperature. The thermal conductivity k is a bulk property of material that indicates its ability to conduct heat. In terrestrial mantles, neither parameter is constant although parameterized thermal evolution models often assume a constant value. In various 2-D and 3-D numerical simulations of mantle convection in the Earth and terrestrial planets, the variation of α and k has been examined (e.g., Ghias and Jarvis, 2008; Hansen et al., 1993; Leitch et al., 1991; Matyska and Yuen, 2005; Schmeling et al., 2003; Tan et al., 2011). The thermal expansivity whose magnitude varies among different mineral phases decreases with pressure and increases with temperature (e.g., Chopelas and Böhler, 1989; Katsura et al., 2009a,b; Xu et al., 2004). Chopelas and Böhler (1989) discovered experimentally that for olivine and magnesium oxides, thermal expansivity decreases with the fifth or sixth power of density up to pressures of about 33 GPa. Theoretical estimates of Anderson (1987) suggest a slower

decrease with the third power of density. The temperature dependence of the thermal expansivity results in an increase of α across the top thermal boundary layer below which α decreases again due increasing pressure (e.g., Tosi et al., 2013c). Since the thermal buoyancy force driving convection and the adiabatic cooling are both proportional to the thermal expansivity, one can expect this variation to have an important influence on the convection pattern in planetary mantles (such as Earth, Venus, or larger terrestrial bodies) where the mantle thickness is not much smaller than the density-scale height. The use of a PT-dependent thermal expansivity lengthens the spatial scale of mantle convection – hot plumes are enhanced as they gain buoyancy while rising through the mantle due to an increase of α with decreasing pressure and the descent of cold downwellings is inhibited as they lose buoyancy while sinking due to a decrease of α with increasing pressure (e.g., Hansen et al., 1993; Leitch et al., 1991; Tan et al., 2011; Tosi et al., 2013c).

Thermal conductivity depends on various factors like temperature, pressure, composition, and texture of the material. In the mantle, the phonon contribution, k_{lat}, and the radiative contribution k_{rad} contribute to the thermal conductivity

$$k_{\mathrm{m}}(T,p) = k_{\mathrm{lat}}(T,p) + k_{\mathrm{rad}}(T) \quad [7]$$

The phonon contribution k_{lat} decreases with increasing temperature T and increases with increasing pressure p, whereas the radiative contribution k_{rad} increases with increasing temperature independent of pressure; for relevant temperatures in a terrestrial mantle, the phonon conductivity is much higher than the radiative contribution. As a consequence, a decrease of the thermal conductivity through the upper part of the mantle, that is, the stagnant lid and the upper thermal boundary layer, and an increase through the mantle beneath are expected. Hofmeister (1999) had developed a model for the temperature and pressure dependence of mantle thermal conductivity under terrestrial conditions:

$$k_{\mathrm{lat}}(T,p) = k_{298}\left(\frac{298}{T}\right)^{a}\exp\left[-\left(4\gamma+\frac{1}{3}\right)\alpha(T-298)\right]\left(1+\frac{K_{\mathrm{T}}'p}{K_{\mathrm{T}}}\right) \quad [8]$$

$$k_{\mathrm{rad}}(T) = 0.01753 - 0.00010365T + 2.2451T^2/10^7 - 3.407T^3/10^{11} \quad [9]$$

where k_{298} is the surface thermal conductivity at 298 K, α is the thermal expansivity and is taken to be constant, γ is the Grüneisen parameter, K_{T} is the bulk modulus, and $K_{\mathrm{T}}' = \mathrm{d}K_{\mathrm{T}}/\mathrm{d}P$. The value of a depends on the chemical bonding, with $a = 0.3$ characteristic of silicates and $a = 0.9$ characteristic of oxides. In parameterized convection models, it is often assumed that the thermal conductivity in the terrestrial mantle has a constant value of 4 W m^{-1} K^{-1} (e.g., Breuer and Spohn, 2003; Hauck et al., 2004; Morschhauser et al., 2011; Schubert et al., 1992; Spohn, 1991; Stevenson et al., 1983). Applying eqns [8] and [9] to terrestrial planets, however, shows that the thermal conductivity is typically lower in the upper mantle and higher for a large part of the lower mantle than the assumed value of 4 W m^{-1} K^{-1}. For instance, in Mars, values may range between 2.5 and 3.5 W m^{-1} K^{-1} at the bottom of the upper thermal boundary layer (Schumacher and Breuer, 2006) and are assumed to be $\sim$7 W m^{-1} K^{-1} at the CMB (Keller and Tackley, 2009; Ruedas et al., 2013a,b). For the Earth, values at the CMB range from 6 W mK^{-1} (de Koker, 2010) to about 10 W mK^{-1} (Goncharov et al., 2008; Ohta et al., 2012; Tang and Dong, 2010) to even 20–30 W mK^{-1} (Hofmeister, 2008).

10.08.3 Planform of Convection

Convection is driven by density variations that arise from thermal and chemical variations. Here, we will be primarily concerned with thermal convection. The Rayleigh number describes the degree of instability and the convective vigor of a convecting system. With increasing convective vigor, thermal boundary layers get thinner and the convection planform smaller (Turcotte and Schubert, 1982). Thus, the horizontal scale of instabilities is reflective of the thickness of the layer itself. In general, the scale of convection depends on the geometry of the system, the heating mode, and the mechanical and thermal boundary conditions. For planetary mantles, shear stress-free conditions with an isothermal top boundary are appropriate, while the heating mode can vary from purely basally heated to purely internally heated with any mixed heating mode in between.

Figure 2 shows the critical Rayleigh number (the Rayleigh number at which convection sets in) for bottom-heated (a) and purely internally heated (b) spherical shells, as a function of the aspect ratio of the shell (η = inner radius/outer radius) for different axisymmetric modes of convection, denoted by spherical harmonic degree l (Zebib et al., 1983). As the shell becomes thinner (η increasing), the harmonic degree of the most unstable mode increases, and the critical Rayleigh number for instability decreases, approaching the plane layer values as η approaches 1. Thus, Mercury's mantle with $\eta = 0.75$ has the lowest critical Rayleigh number of all the terrestrial planets. Nonetheless, Mercury's small size leads to a low intrinsic Rayleigh number for the shell and it is not clear if convection occurs at all (e.g., Hauck et al., 2004; Michel et al., 2013; Tosi et al., 2013a).

For small core sizes (low η), the most unstable modes have very low l, and for η below about 0.25, degree-1 mode is the most unstable. This mode has a single, cylindrical upwelling and a broader, weaker downwelling located antipodally. It was thought for some time that this convection mode described the Martian interior, with the huge volcanic upland of Tharsis located above the upwelling. The measurement of the moment of inertia of Mars (Folkner et al., 1997), however, found that the aspect ratio of the Martian mantle is much larger than 0.25, and other solutions for a degree-1 mode were found involving phase transitions (see succeeding text and **Section 10.08.6.4**). Higher-degree modes have more upwellings and downwellings, taking the form of either sheets or columns depending on the polarity of the solution. Sheetlike downwellings seem to be the preferred geometry for planetary mantles, but upwellings may also organize along linear structures (Schubert et al., 2001), perhaps explaining the chains of volcanic highlands such as those seen on Venus (Ratcliff et al., 1997) or the line of large volcanoes in the Tharsis region of Mars.

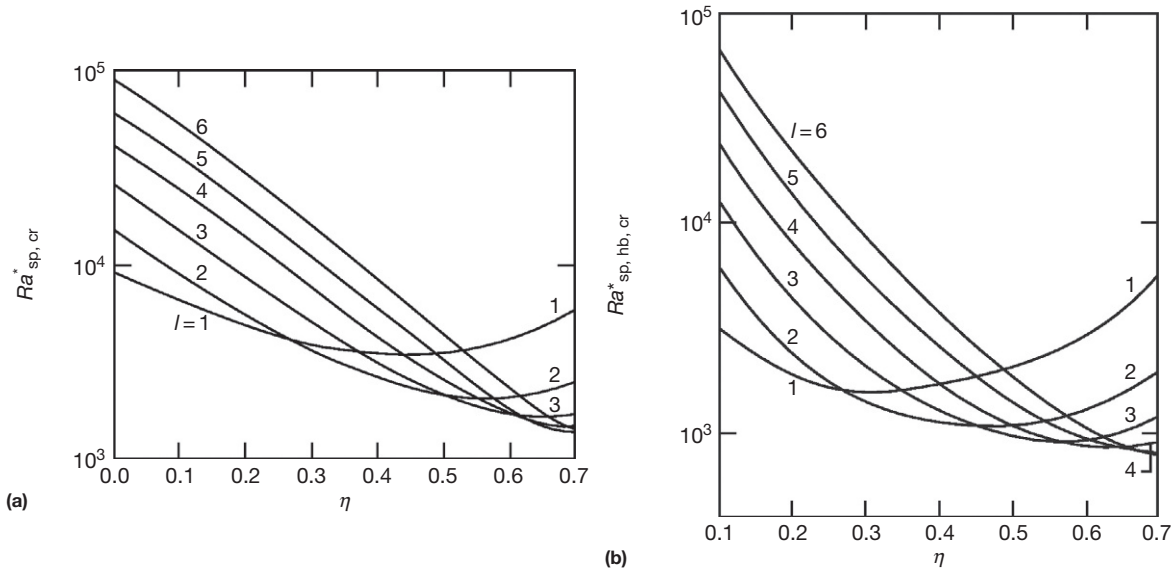


Figure 2 Critical Rayleigh number for bottom-heated (a) and purely internally heated (b) spherical shells, as a function of the aspect ratio η of the shell. The curves for different modes are labeled by spherical harmonic degree l. Reproduced from Zebib A, Schubert G, Dein JL, and Paliwal RC (1983) Character and stability of axisymmetric thermal convection in spheres and spherical shells. *Geophysical and Astrophysical Fluid Dynamics* 23: 1–42. (http://www.informaworld.com).

For a given shell aspect ratio, increasing the Rayleigh number above the critical value causes additional modes to become unstable, some larger-scale (smaller l) and some smaller-scale. As the Rayleigh number continues to increase, more and more high-degree modes become unstable, and the scale of these modes becomes small compared with the depth of the layer. At this point, the cellular nature of the convection breaks down, and the flow is better described by the transient formation of small-scale instabilities in the boundary layers. This happens for Rayleigh numbers of 10^7–10^8, which is close to the values expected for the mantles of Venus and Earth. It has been suggested that the volcanic coronae unique to Venus are the surface expression of such instabilities (Stofan et al., 1991).

Another factor that influences the planform of convection is the presence of mantle phase changes. In Venus and Earth, at least two significant solid-state phase changes associated with the transformation of olivine-to-spinel and spinel-to-perovskite structure are present. The olivine-to-spinel transition is exothermic and has a positive Clapeyron slope and thus can act to enhance convective motions, while the spinel-to-perovskite phase change is endothermic and has a negative slope and thus can act to retard convection (Schubert et al., 1970). The influence of phase changes on convection planform has two aspects: First, if the endothermic phase change can act as a significant barrier to convection, the layer may be split into two independently (or nearly so) convecting layers, each with a higher aspect ratio and a lower Ra than the whole mantle. Second, the effectiveness of the endothermic phase change in resisting convective motions is wavelength-dependent, becoming less effective as the wavelength increases (Bercovici et al., 1993). Thus, this transition can act as a filter, preventing short-wavelength instabilities from passing but allowing larger-scale motions (Tackley, 1996; Tackley et al., 1993, 1994). For Venus, this perhaps explains the large-scale structure of the highland regions superimposed on the small-scale features that generate the coronae. In Mars, the olivine-to-spinel transition falls near the middle of the mantle, but the spinel-to-perovskite transition, if it is reached at all, is located very near the core. Convection calculations with this geometry show that the number of upwellings originating from the CMB is reduced by the presence of the endothermic phase transition and that single-plume planforms are possible (Harder, 1998; Harder and Christensen, 1996).

Convection planform is also influenced by rheological stratification. Although we know very little of the interior structure of the other terrestrial planets, the Earth shows a strong rheological contrast between the upper mantle and lower mantle, with estimates ranging from a factor of 30 to 1000 in the viscosity difference between the upper mantle and lower mantle (e.g., Cadek and Fleitout, 2003; Hager and Richards, 1989; Kaufmann and Lambeck, 2000; Peltier and Jiang, 1996). Additionally, the viscosity of mantle materials shows a continuous increase with pressure. Finally, the Earth exhibits a global or nearly global low-viscosity zone (asthenosphere) beneath the lithosphere. The rheological stratification of the Earth affects the length scales of convection in two ways: First, the high-viscosity lower mantle has an effective Rayleigh number that is lower than the mantle as a whole, and thus, the scale of instabilities will increase relative to the upper mantle. Second, the low-viscosity zone acts to decrease the dissipation associated with shear across the upper thermal boundary layer, allowing longer cells to form (Busse et al., 2006; Lenardic et al., 2006). It is not clear if either of these effects is present in the other terrestrial planets. The low-viscosity zone, in particular, is thought to be related to the recycling of water into the mantle by subduction, a process that is at present not relevant to the other bodies.

10.08.4 Thermal Evolution Models Using Parameterized Convection

Thermal evolution calculations with 2-D or 3-D mantle convection codes are very time-consuming on present-day computers. In these calculations, a set of differential equations have to be solved for the conservation of mass, momentum, and energy including an appropriate constitutive equation. Because of the inherent complexity in these models, it is often desirable to take an empirical approach and parameterize the convective heat transfer rate as a function of known quantities. Such parameterizations can be derived using simple theories, which result in scaling laws that describe the heat transport in the interior as a function of the convective parameters.

Our improving understanding of the heat transport mechanisms on terrestrial planets over the last two decades has led to repeated changes in the preferred scaling law used to model the thermal evolution of one-plate planets. Initially, the scaling law for a fluid with constant viscosity was used for one-plate planets. In the meantime, it has been recognized that this scaling law models the heat transport in a planet where convection comprises the whole mantle including the outer layers. In fact, such a model describes the heat transport in a planet with plate tectonics better than that in a one-plate planet (e.g., Schubert et al., 2001). The plate tectonic regime is expected to cool the planet very efficiently because the comparatively cold outer layers become recycled into the interior of the mantle by convection (Figure 3).

Subsequent attempts to model the heat transport in one-plate planets included the effects of a growing lithosphere in the models with parameterizations based on constant viscosity scaling laws (e.g., Schubert et al., 1990). The base of the non-convecting lithosphere is represented by an isotherm characteristic of the transition from viscous deformation to rigid response to loads applied over geologic timescales (e.g., Schubert et al., 1992). Such a model represents the heat transport in a planet with a single plate on top of a convecting mantle. However, this model assumes that the lid coincides with the rheological lithosphere, that is, the region of the planet's crust and mantle over which the flow properties of rock are insufficient to relieve stress. The part of the upper mantle that is weaker than the rheological lithosphere is assumed to be constantly recycled within the mantle. In comparison to the recently derived stagnant lid parameterization, these models, therefore, represent a mechanism of lithosphere delamination. The efficiency of the heat transport by lithosphere delamination is in between that of plate tectonics and that of stagnant lid convection.

More recently, new scaling laws have been derived from convection models in layers of fluids with strongly temperature-dependent viscosities (e.g., Davaille and Jaupart, 1993; Deschamps and Sotin, 2000; Grasset and Parmentier, 1998; Hüttig and Breuer, 2011; Moresi and Solomatov, 1995; Reese et al., 1999; Richter, 1978; Solomatov, 1995; Solomatov and Moresi, 2000). These scaling laws have been suggested to represent the heat transport in a planet with a single plate on top of a convecting mantle. This stagnant lid convection predominantly cools the outer layers of a planet through stagnant lid growth while the deep interior is cooled rather inefficiently.

In the following, various approaches to model the heat transport that are based on either constant (Section 10.08.4.1) or temperature-dependent (Section 10.08.4.2) viscosity fluids will be introduced.

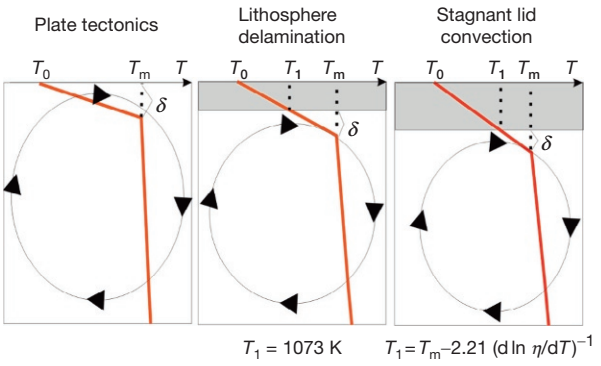


Figure 3 Sketch of the convective flow (circle with arrows) and the temperature profiles (red lines) in a mantle for three different heat transport mechanisms. In the plate tectonic regime, mantle convection comprises the entire mantle including the outer layers. For this regime, the temperature difference across the thermal boundary layer (δ), which drives the convection, is given by $T_m - T_0$ (with T_m the interior temperature of the convecting mantle below the thermal boundary layer and T_0 the surface temperature). In the stagnant lid model and the lithosphere delamination model, mantle convection occurs below a lid (shaded area). The temperature difference across the thermal boundary layer is given by $T_m - T_l$ (with T_l the temperature at the base of the lid). T_l differs between these two models, being constant for the lithosphere delamination model and dependent on T_m and the mantle rheology for the stagnant lid model.

10.08.4.1 Parameterizations Based on Constant Viscosity Laws

From convection experiments and boundary layer theory for constant viscosity fluids, the classical power law relationship between heat flow and convective parameters (e.g., Roberts, 1979; Turcotte and Oxburgh, 1967) has been derived:

$$Nu = aRa^{\beta} \qquad [10]$$

This equation relates the dimensionless heat flux out of the convecting layer, expressed as the Nusselt number, Nu, to the strength of thermal convection as measured by the Rayleigh number, Ra. The parameters a and β are constants, with a typically ranging between 0.195 and 0.339 and β between 1/4 and 1/3 depending on the geometry of the convecting layer, the heating mode, and the boundary conditions (e.g., Deschamps and Sotin, 2000; Deschamps et al., 2010; Jarvis, 1984; Ratcliff et al., 1996; Schubert and Anderson, 1985; Sotin and Labrosse, 1999; Wolstencroft et al., 2009; Yanagisawa and Yamagishi, 2005; Zebib et al., 1985; Zhong, 2005).

The Nusselt number is a measure of the ability of convection to transfer heat in comparison with thermal conduction. It is calculated by dividing the average surface heat flux, q_s, by the theoretical conductive heat flux, $k\Delta T_d/d$, where d is the

thickness of a conductive fluid layer, k is the thermal conductivity, and ΔT_d is the temperature jump across the fluid layer:

$$Nu = q_s \frac{d}{k\Delta T_d} \quad [11]$$

For convection, the surface heat flow is given by

$$q_s = k\frac{\Delta T_\delta}{\delta} \quad [12]$$

where δ is the thickness of the upper thermal boundary layer and ΔT_δ is the temperature difference across this boundary layer. With convection, Nu is larger than one. The Rayleigh number is defined as

$$Ra = \frac{\alpha\rho g\Delta T_d d^3}{\kappa\eta} \quad [13]$$

where α is the thermal expansivity, ρ is the density, κ is the thermal diffusivity, g is the acceleration due to gravity, and η is the viscosity of the fluid.

The power law in eqn [10] has been applied to model the thermal evolution not only of terrestrial planets, for example, of the Earth with a plate tectonic regime, but also of one-plate planets such as Mercury, Venus, and Mars (e.g., Schubert, 1979; Schubert et al., 1979a; Sharpe and Peltier, 1978; Stevenson et al., 1983). Equation [10] has been used although the viscosity η of the terrestrial mantles is thought to be strongly temperature-dependent (eqn [6]).

The models for one-plate planets have later been modified to include the effects of a growing lithosphere by Spohn and Schubert (1983) (see also Schubert et al., 1992; Spohn, 1991) using an equation for the growth of a lithosphere, which was first proposed by Schubert et al. (1979b):

$$q_m - q_l = \rho_m C_m (T_m - T_l)\frac{dl}{dt} \quad [14]$$

where q_m and q_l denote the heat flow into the lithosphere from the convecting mantle and the heat flow through the lithosphere by conduction, respectively; C_m is the heat capacity; T_l is the temperature at the base of the lithosphere; T_m is the temperature in the mantle at the base of the upper thermal boundary layer; l is the thickness of the lithosphere; and t is the time. The base of the lithosphere in these models is an isotherm assumed to be characteristic for the transition from viscous deformation to rigid response to loads applied over geologic timescales. This lithosphere can be identified with a rheological lithosphere. A representative value of T_l is around 1073 K.

10.08.4.2 Stagnant Lid Parameterization Based on Temperature-Dependent Viscosity Laws

In recent years, experimental and numerical studies (e.g., Davaille and Jaupart, 1993; Deschamps and Sotin, 2000; Grasset and Parmentier, 1998; Hüttig and Breuer, 2011; Moresi and Solomatov, 1995; Richter, 1978; Reese et al., 1998; Reese et al., 1999; Solomatov, 1995; Solomatov and Moresi, 2000) have improved our understanding of convection in layers of fluids with strongly temperature-dependent viscosities. It has been shown that the simple scaling law (eqn [10]) is only applicable if the viscosity ratio across the layer is relatively small.

Three basic regimes of convection are found to occur, depending on the variation in viscosity between the hot interior of the convecting fluid and the cold surface: the small-viscosity regime, the transitional regime, and the stagnant lid regime. If the viscosity ratio across the layer is relatively small (smaller than about 100), the cold thermal boundary layer is mechanically similar to the rest of the fluid, and the convection planform and the temperature distribution are essentially the same as in constant viscosity convection. For larger viscosity ratios (between about 100 and several thousand), the cold thermal boundary layer participates in the convection but deforms more slowly than the interior of the convecting fluid and thereby controls the heat transfer (transitional regime). With a viscosity ratio greater than about 10^4, a so-called stagnant lid will form on top of the underlying convecting layer. This critical viscosity ratio also depends on the increase of the viscosity with depth (Hüttig and Breuer, 2011). This stagnant lid (**Figure 3**) is not participating in the convective flow and most of the viscosity variation is within the lid.

Convection in the layer below the stagnant lid is driven by the nonadiabatic temperature difference across it. The viscosity variation in this convecting layer is small, that is, by only about one order of magnitude. Numerous studies assuming heating from below (e.g., Moresi and Solomatov, 1995; Morris and Canright, 1984; Solomatov, 1995) or assuming heating from within (e.g., Choblet and Sotin, 2000; Grasset and Parmentier, 1998; Hüttig and Breuer, 2011; Solomatov and Moresi, 2000) find a viscosity ratio in the convecting layer smaller than 10. This observation is also independent of the assumed mantle rheology law, whether or not it is linear (Newtonian) or a power law (non-Newtonian).

All terrestrial planets other than the Earth presently show a stagnant lid on their surface. For such a lid to form, the following condition is required (Solomatov, 1995):

$$\theta \equiv \frac{d\ln\eta}{dT}\Delta T > \ln 10 \quad [15]$$

For commonly assumed mantle rheology parameters, including the activation energy with $A \sim 400$ kJ mol^{-1} (e.g., Turcotte, 1982), and typical deep mantle temperatures of terrestrial planets (i.e., between about 1600 and 2500 K), the parameter θ is much larger than 10, which would imply that all terrestrial planets would be in the stagnant lid regime. A problem with the stagnant lid model, however, is that it may be ignoring some important aspects of mantle convection since the Earth also meets the previously mentioned criterion for stagnant lid convection but instead has plate tectonics. For plate tectonics to occur, more complex rheologies need to be considered that mimic, for instance, distributed brittle deformation (e.g., Bercovici, 2003; Höink et al., 2012; Moresi and Solomatov, 1998; Tackley, 2000; Trompert and Hansen, 1989) because simple temperature-dependent rheologies lack the modes of deformation responsible for the lithosphere strain localization and weakening responsible for plate boundary formation. Of course, such deformation mechanisms must be plausibly inoperative or ineffective on the other terrestrial planets lacking plate tectonics.

Davaille and Jaupart (1993) had shown with their experiments on internally heated fluids that the temperature at the base of the stagnant lid, T_l, can be expressed as a function of the temperature of the underlying convecting layer, T_m, measured just below the thermal boundary layer, and the rate of change of viscosity with temperature:

$$T_l = T_m - 2.21\left(\frac{d\ln\eta}{dT}\right)^{-1} \quad [16]$$

Assuming a mantle viscosity as described by eqn [6] without pressure dependence, for stagnant lid convection, the temperature difference across the upper thermal boundary layer that drives the convection of the fluid (see Figure 3), if heated from within, is

$$\Delta T_{SL} = T_m - T_l = 2.21\frac{RT_m^2}{A} \quad [17]$$

A similar relation has been obtained from numerical experiments by Grasset and Parmentier (1998) also with internally heated fluids. Choosing representative parameter values for a terrestrial mantle with $T_m \sim 1800$ K and $A \sim 400$ kJ mol^{-1}, the difference between T_m and T_l is small, of the order of 100 K. This temperature difference across the thermal boundary layer is small in comparison to $(T_m - T_l)$ in earlier models based on constant viscosity laws. In these models, T_l equals the surface temperature and $T_m - T_l$ is about 1200 K or, if a rheological lithosphere is considered, T_l is about 1073 K (e.g., Schubert et al., 1979b). In that case, $T_m - T_l$ is about 700 K. Thus, the stagnant lid comprises the rheological lithosphere and is likely to be much thicker than the latter. As a consequence of the relatively high temperatures at the base of the stagnant lid, material in the lower part of the lid may have comparatively low viscosities and may be easily deformable on geologic timescales.

In the present literature, there are at least two different approaches to parameterize the heat transport through a strongly temperature-dependent fluid in the stagnant lid regime. These two approaches differ in the time-dependent versus quasi-equilibrium treatment of the stagnant lid and are otherwise equivalent:

1. The approach by Grasset and Parmentier (1998) is based on the following rules for the stagnant lid regime: The interior of a layer adjusts so that the viscosity contrast across the convective thermal boundary layers is about 10. In the convecting layer, the scaling law for a fluid with constant viscosity is valid, and the Nusselt–Rayleigh number relationship for a fluid with constant viscosity is used (eqn [10]). The thickness of the stagnant lid is determined by equating the heat flux through the lid to the heat flux from the underlying convecting layer into the lid, according to eqn [14]. Instead of using an isotherm at the base of the stagnant lid, the temperature is time-dependent and calculated from eqn [16]. The conductive heat transfer rate within the stagnant lid is calculated from a time-dependent heat conduction equation.
2. In the second approach (e.g., Moresi and Solomatov, 1995; Reese et al., 1999; Solomatov, 1995; Solomatov and Moresi, 2000), a scaling law is derived for the whole layer, that is, the convecting part of the layer and the stagnant lid:

$$Nu = a\theta^c Ra_i^\beta \quad [18]$$

The constants a, c, and β vary with the properties of the convective regime. The internal Rayleigh number, Ra_i, is based on the total temperature drop across the layer and on the viscosity evaluated at the temperature of the interior below the stagnant lid and the upper thermal boundary layer. In addition to the internal Rayleigh number, a second parameter is introduced for this approach, the Frank–Kamenetskii parameter, θ, which is the natural logarithm of the viscosity contrast across the convecting layer (eqn [15]).

10.08.4.3 Plate Tectonics, Lithosphere Delamination, and Stagnant Lid Convection

The parameterizations based on constant and temperature-dependent viscosity laws introduced in Sections 10.08.4.1 and 10.08.4.2 are used to describe the heat transport mechanisms in a terrestrial planet by either plate tectonics, lithosphere delamination, or stagnant lid convection (Figures 3 and 4). In the following, we use for all model calculations parameters that correspond to the planet Mars (see Table 1).

To calculate the thermal evolution, the energy balance of the mantle and the core has been solved:

$$\rho_m C_m V_m \frac{dT_m}{dt} = -q_m A_m + q_c A_c + Q_m V_m \quad [19]$$

$$\rho_c C_c V_c \frac{dT_c}{dt} = -q_c A_c \quad [20]$$

where V_m and V_c are the volume of the convecting mantle and the core; A_m and A_c are the surface area of the mantle and the core; T_m and T_c are the temperature of the isothermal mantle below the conductive layer and the temperature of the core, respectively; q_m and q_c are the heat flow out of the mantle and the core; and Q_m is the heat production rate in the mantle

$$Q_m = Q_0 \exp(-\lambda t) \quad [21]$$

where Q_0 is the initial heat generation rate, λ is the decay constant, and t is the time.

The heat flow of the mantle and the core is calculated from eqn [12] with δ the thickness of the upper or lower thermal boundary layer and ΔT_δ the temperature difference across the upper or lower boundary layer. The thickness of the boundary layers is calculated from eqn [13] replacing d with δ and assuming that the layers become unstable when Ra reaches a critical value, Ra_{crit}. The lid of the stagnant lid and the lithosphere delamination model grow in thickness with time according to eqn [14].

Stagnant lid parameterization models show that the lid will thicken rapidly as the planet cools, while the temperature of the underlying convecting mantle and core will change comparatively little (Figure 4(a)–4(c)). With plate tectonics, there

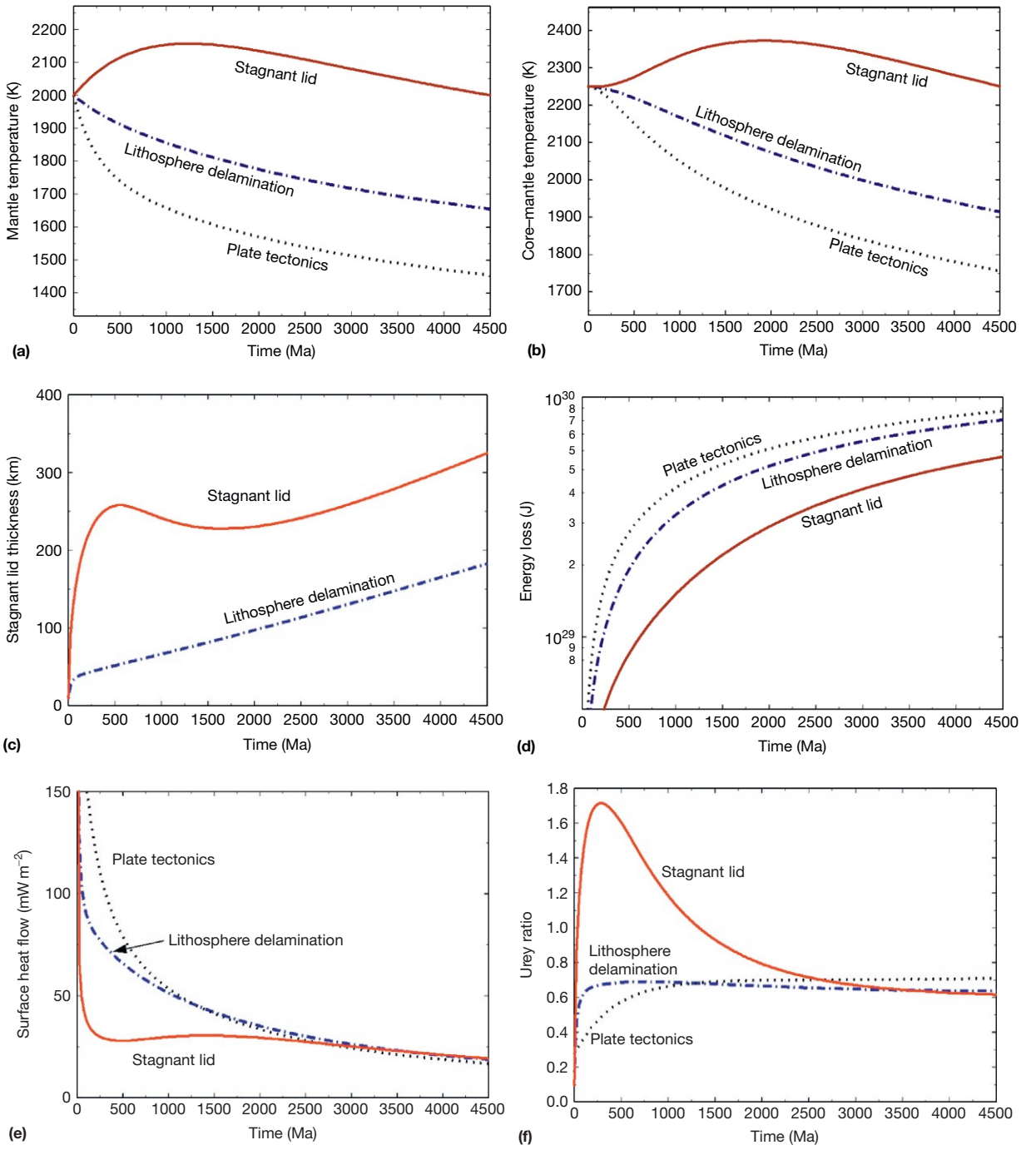


Figure 4 Mantle temperature (a), core–mantle temperature (b), stagnant lid thickness (c), energy loss (d), surface heat flow (e), and Urey ratio (f) as a function of time for three different heat transfer models: stagnant lid convection (red line), lithosphere delamination (dash-dotted blue line), and plate tectonics (dotted line). For further description, see text.

is, of course, no growth of a lid and the cooling of the deep interior is most efficient. The evolution of the lid thickness and the mantle temperature for the lithosphere delamination model is in between the two extremes. For the stagnant lid model, the conductive layer grows rapidly during the first few hundred million years and its thickness varies comparatively little thereafter. Plate tectonics and lithosphere delamination models show an approximately constant rate of increase in the conductive layer thickness. Although the present-day thickness of the stagnant lid differs significantly among the models (320 km for the stagnant lid model, 180 km for the lithosphere delamination model, and no lid for the plate tectonic model), the present-day thickness of the conductive layer is relatively similar for all three models.

Table 1 Definition of parameters and their values for the thermal evolution models in **Section 10.08.4.3**

Parameter	*Notation*	*Value*	*Unit*
Radius of the planet	R_p	3400×10^3	m
Radius of the core	R_c	1700×10^3	m
Gravity	g	3.7	$m\ s^{-2}$
Density of the mantle	ρ_m	3500	$kg\ m^{-3}$
Heat capacity of the mantle	C_m	1000	$J\ kg^{-1}\ K^{-1}$
Mantle thermal expansivity	α	2×10^{-5}	K^{-1}
Mantle thermal diffusivity	κ	10^{-6}	$m^2\ s^{-1}$
Mantle thermal conductivity	k	4	$W\ m^{-1}\ K^{-1}$
Surface temperature	T_0	220	K
Gas constant	R	8.3144	$J\ mol^{-1}\ K^{-1}$
Initial heat source density	Q_0	1.6×10^{-8}	$W\ m^{-3}$
Decay constant	λ	10^{-17}	s^{-1}
Exponent of the Nu–Ra relation	β	1/3	
Critical Rayleigh number	Ra_{crit}	450	

Figure 4(d) and 4(e) compares the cumulative energy loss, that is, the energy loss over the entire evolution of the planet and the surface heat flow of the three models with $A = 300\ kJ\ mol^{-1}$. Stagnant lid convection cools the entire planet least effectively. After 4.5 Ga, the energy loss is about 3.1×10^{29} J smaller for the stagnant lid model compared to the plate tectonic model and 2.3×10^{29} J smaller for the stagnant lid model compared to the lithosphere delamination model. Such energy losses, when distributed homogeneously over the entire mantle, imply that the present-day mantle is about 600 K cooler in the plate tectonic model and about 450 K cooler in the lithosphere delamination model than it is in the stagnant lid model.

It may be speculated that a planet, in which heat is transported effectively by plate tectonics, has a higher surface heat flow than a planet in the stagnant lid regime. However, this is only true in the early part of its evolution. After about 2.5 Ga, the surface heat flow becomes slightly higher in the stagnant lid model. The albeit small difference in the present-day surface heat flow has also implications for the Urey ratio, that is, the ratio of the heat produced within the mantle by radioactive decay to the total surface heat flow (Figure 4(f)). The present-day Urey ratio for our Mars-like test planet is about 0.6 for the stagnant lid regime and 0.7 with plate tectonics. For the lithosphere delamination model, the present-day value is in between the two models of stagnant lid convection and plate tectonics. The contribution of secular cooling at present is, therefore, slightly higher for a planet in the stagnant lid regime than for a planet with plate tectonics assuming the same mantle parameters.

A strong decrease of the surface heat flow can be observed during the first 250 Ma for the stagnant lid model. The strong decrease in the surface heat flow is caused by the adjustment of the lithosphere, which grows rapidly from 10 to 250 km as a consequence of the inefficient heat flow out of the convecting mantle and the more efficient heat transport through the conductive layer. The thicker the assumed initial lithosphere and the lower the initial mantle temperatures, the slower is the decrease of the surface heat flow in the early evolution. The cooling behavior during the early stage of the planet's evolution is, therefore, strongly dependent on the initial conditions. It has been argued by Choblet and Sotin (2000) that an initial stage up to 1 Ga cannot be described by the stagnant lid parameterization, which is based on steady-state convection models. During that initial transitional stage, the mantle cools by conduction until the upper thermal boundary layer becomes unstable due to increased mantle temperatures and reduced viscosities, and convection sets in. The length of this initial stage depends strongly on the early temperature distribution, that is, the hotter the early planet, the shorter is this transitional phase (Choblet and Sotin, 2000). As in general a high interior temperature up the melting temperature in the upper mantle is suggested for terrestrial planets (see **Section 10.08.2.2**), the stagnant lid parameterization can be applied after a few tens to hundred million years after planet formation.

10.08.5 Thermal Evolution, Volcanic History, and Magnetic Field History of Terrestrial Planets

The various parameterized approaches, described earlier in the text, have been used to calculate the thermal histories of the terrestrial planets and a significant amount of work has been published to date (e.g., Breuer and Spohn, 2003; Fraeman and Korenaga, 2010; Grott et al., 2011a,b; Hauck and Phillips, 2002; Morschhauser et al., 2011; Schubert, 1979; Schubert et al., 1979a, 1990, 1992; Sharpe and Peltier, 1978; Spohn, 1991; Stevenson et al., 1983). According to these previous models, the terrestrial planets are differentiated rapidly into a solid mantle and a fluid core (see also **Sections 10.08.2.2.1** and **10.08.2.2.2**). The temperature distribution subsequent to core formation, therefore, serves as the initial boundary condition for the thermal evolution models. Early thermal evolution models use fixed initial temperatures (e.g., Breuer et al., 1993; Schubert and Spohn, 1990; Schubert et al., 1992; Spohn, 1991). The main reason not to vary the initial temperature distribution, which is basically unknown, can be justified by a characteristic feature of the thermal evolution models: The strong temperature dependence of mantle viscosity acts as a thermostat regulating the average mantle temperature (Tozer, 1965). For a hot planet, mantle viscosity is low, and extremely vigorous convection rapidly cools the planet. For a relatively cool planet, the mantle viscosity is higher, and more moderate convection cools the planet at a reduced rate. Self-regulation tends to quickly bring the viscosity of the mantle to a value that facilitates efficient removal by convection of the heat generated in the mantle. Even an initially cold mantle would heat up by radioactivity until the self-regulated viscosity is reached. As a consequence of the self-regulation, the present state of the convecting mantle has little or no memory of the initial conditions. This circumstance makes the early thermal evolution models more applicable, in particular because the initial state is unknown. However, if one is interested in crustal growth, atmosphere evolution, and magnetic field generation of a planet, in particular in its early evolution, knowledge about the early temperature distribution is required

(e.g., Breuer and Spohn, 2003, 2006; Fraeman and Korenaga, 2010; Grott et al., 2011a,b; Hauck and Phillips, 2002; Morschhauser et al., 2011).

Crustal growth and dynamo action in the core are strongly coupled with the thermal evolution of a planet since both processes depend on the vigor and the mode of convection in the mantle. A model of a planet's evolution and tectonics is therefore required to explain the evolution of the magnetic field, and it must also account for the formation of the core of the planet and for the formation and growth of its crust. Furthermore, the thickness of the elastic lithosphere places important constraints on the thermal evolution of a planet. The elastic lithosphere is the upper layer of a terrestrial planet that supports stresses over geologically long ($\sim 10^8$ a) intervals and its thickness can be derived by a joint analysis of the gravity and topography data or from flexural modeling to match the topography (see **Chapter 10.05**). Since the base of the elastic (mechanical) lithosphere can be defined by an isotherm that depends on the rheology (composition) and the strain rate (McNutt, 1984), its thickness allows rough estimates of the surface heat flow. It is important to note that the elastic lithosphere thickness obtained by gravity and topography data does not represent the current state of the thermal lithosphere. It rather shows the thermal state at the time of formation of the considered geologic structure. Thus, the elastic lithosphere thickness provides constraints on models of thermal evolution.

For a typical terrestrial mantle material dominated by olivine, this isotherm can be estimated to be about 1050 K using the rheology parameters of Karato et al. (1986) and a strain rate of 10^{-17} s^{-1}. This isotherm defines the maximum thickness of the elastic lithosphere since the thickness of the effective elastic lithosphere decreases with its deflection. Furthermore, the crust can significantly reduce the elastic thickness since the crust is intrinsically weaker than the mantle (e.g., Burov and Diament, 1995).

10.08.5.1 Crustal Formation

The crust of a terrestrial planet forms when partial melt of the mantle rises toward the surface. Thus, the composition, the thickness, and the time of emplacement of the crust can give us important constraints on the thermochemical evolution; it tells us about the interior temperature distribution and how the processes forming the crust vary with time. On the other hand, the crust itself influences the thermal evolution. (1) Melt that forms the crust is enriched in radioactive elements and in volatiles such as water and CO_2. The consequence is the redistribution of the radioactive elements and the volatiles and in particular a depletion of the mantle in these elements. (2) Crustal material has a different composition and structure than mantle material. The consequence is a lower thermal conductivity of the crust in comparison to the conductivity of the mantle, thus a less efficient heat transport through the outer layer.

10.08.5.1.1 Primary and secondary crustal formation

The process of crustal formation on terrestrial planets can, in general, be divided into two phases: primary crustal formation and secondary crustal formation. A primary crust can be formed as a consequence of efficient differentiation due to core formation and/or freezing of a magma ocean. Such a primordial crust can be strongly enriched in radioactive elements as suggested for the anorthositic crust on the Moon and may have strongly depleted the mantle (for reviews, see Warren, 1985, and Elkins-Tanton, 2012). Except for the Moon, the existence, persistence, and composition of such a crust for the terrestrial planets are however unknown.

The presence of an early magma ocean on a terrestrial planet suggests chemical stratification of the mantle due to fractional crystallization. Fractional crystallization of a cooling magma ocean is generally accepted as an explanation of the observed dichotomy between the Eu-enriched anorthositic crust and the Eu-depleted basaltic source regions of the Moon (e.g., Wood et al., 1970). Assuming the magma ocean freezes rapidly during about 100–200 Ma (e.g., Elkins-Tanton et al., 2003, 2005; Minear and Fletcher, 1978; Solomon and Longhi, 1977) without any disturbances, this part of the mantle most likely became chemically stratified following magma ocean differentiation and crystallization with the late dense iron-rich phases coming to rest upon a less dense Mg-rich phase. Furthermore, most of the incompatible and heat-producing elements would have been enriched in the residual melt forming the so-called KREEP-rich (standing for potassium, rare earth elements, and phosphorus) layer and leaving the mantle of the original magma ocean strongly depleted in radioactive heat sources. In case of the Moon, this layer is located just below the anorthositic (plagioclase-rich) crust that formed through flotation of the low-density plagioclase to the top of the magma ocean. On larger planets such as Mars and the Earth, plagioclase does not become stable due to the higher-pressure gradient until the magma ocean has solidified to a high degree close to the planetary surface, where high crystal fractions also prevent flotation. Furthermore, water in silicate melts suppresses plagioclase (e.g., Goldsmith, 1982). Thus, for large and wet planets assuming a global and fractional crystallizing magma ocean, a dense crust highly enriched in iron-rich phases and radioactive elements is more likely (Elkins-Tanton et al., 2003).

Chemical fractionation, which results in mantle layers of different compositions and densities, strongly influences the subsequent thermochemical evolution of the planet (e.g., Elkins-Tanton et al., 2003; Hess and Parmentier, 1995). The unstable layering due to fractional crystallization (dense cumulate material on top of a less dense material below) may result in a global overturn of the crustal and mantle material. However, for the larger terrestrial planets like Earth and Venus, and possibly for Mars, Tonks and Melosh (1990) argued that the higher-pressure gradient for those planets as compared to the Moon and Mercury inhibits efficient chemical fractionation. The magma ocean rather would have started to cool via equilibrium crystallization (Solomatov and Stevenson, 1993). The existence of a magma ocean in general has been questioned from numerical modeling of a growing Mars (Senshu et al., 2002). The models show no global magma ocean formed during accretion; instead, small magma ponds form by the impacts – these models, however, neglect, for instance, the efficient heating by short-lived radioactive isotopes and assume that Mars formed only from small planetesimals.

In the subsequent evolution after freezing of the magma ocean, a secondary crust that is usually basaltic is formed by

partial melting of the mantle. Depending on the degree of early mantle differentiation, the entire mantle (or only part of the upper mantle) is already depleted in crust-forming material. The formation of the secondary crust depletes the mantle further in radioactive elements as those are incompatible and enriched in the melt that form the crust. The mechanisms of crustal formation differ between a plate tectonics and a one-plate planet.

Plate tectonics provides for crustal formation in a two-step fashion (see Condie (1997) for a discussion of plate tectonics and the Earth's crust and Schubert et al. (2001) for a discussion of plate tectonics and mantle convection). Crustal formation is most effective at divergent plate boundaries, where rising hot mantle material crosses the solidus near the surface. This pressure-release melting generates basaltic crust that is continuously recycled at convergent plate boundaries. More silicic crust is produced in a second differentiation step at convergent plate boundaries where basaltic crust is remelted (together with continental sediments and possibly mantle rock) to form new continental-like crust. Efficient crustal recycling – in particular of the oceanic crust – occurs at the subduction zones at plate margins.

For one-plate planets, there is no two-stage differentiation. Instead, melt is formed underneath the lithosphere usually at greater depth than with plate tectonics. The crust produced is assumed to be basaltic. In contrast to crustal formation at the divergent plate boundaries on plate tectonic planets, crustal growth is limited by the increasing thickness of the lithosphere as the planet cools. First, because buoyancy of the melt decreases with increasing depth to the source region, the melt is more compressible than the mantle rock. Second, the melt needs to rise through the ductile lower and the upper elastic parts of the lithosphere. For the latter, the permeability of the elastic part of the lithosphere to magma ascent decreases with increasing thickness and pressure. Nonetheless, volcanism is observed even on a body with a thick crust and weak gravity such as the Moon. Crustal recycling is not efficient in one-plate planets and is associated with crustal delamination that requires the lowermost crust to undergo mineral equilibration under eclogite facies. The lower crustal layer then becomes denser that the underlying mantle (e.g., Arndt and Goldstein, 1989) and is able to sink into the interior. The basalt–eclogite transition occurs for Earth at a pressure of about ~1.2 GPa (45 km on Earth and 50 km on Venus) and a temperature higher than ~850 K. A different composition further influences the occurrence of the transition. For example, the basalt–eclogite transition on Mars would occur at a lower pressure than for the Earth, due to its iron-rich composition (Babeyko and Zharkov, 2000). Secondary crustal formation may also result in a chemical layering of the mantle as suggested for the primary crust that forms from freezing of a magma ocean. When basaltic melt is extracted from the mantle and segregates to form the surface layer, fertile mantle rock reduces to harzburgite. Under certain melting conditions (Schutt and Lesher, 2006), harzburgite is less dense than its fertile source rock, and this layer can have a stabilizing effect on the mantle dynamics and prevent further cooling of the planet. Depending on the thickness of the depleted layer (e.g., Fraeman and Korenaga, 2010; Ogawa and Yanagisawa, 2011; Plesa and Breuer, 2014; Ruedas et al., 2013a; Schott et al., 2001), three different evolution scenarios are possible: (1) The layer is located in and thinner than the stagnant lid and thus has no influence on the thermal evolution of the planet. (2) The harzburgite layer grows faster than the stagnant lid, further slowing the cooling of the planet as the thickness of the nonconvecting part of the mantle increases and inefficiently transports heat by conduction. (3) The harzburgite layer is sufficiently thick that it can convect separately from the lower fertile mantle. Such two-layer convection results in inefficient cooling of the deep interior. The thickness and the evolution of the depleted harzburgite layer depend on the mantle composition associated with the density difference between primordial mantle and depleted mantle, on the mantle temperature, and especially on the depth of melting, that is, for high temperature, the melting is deeper, and the depleted layer is thicker.

10.08.5.1.2 Plume volcanism and partial melt zone

As mentioned in the preceding text, in one-plate planets, melt is usually generated underneath the lithosphere in places where the mantle temperature is higher than the mantle solidus. The melt can be produced either locally due to upwelling hot material, that is, mantle plumes, or more globally if the average mantle temperature is close to the solidus.

Whether a plume and associated volcanism can occur depends on the existence of a thermal boundary layer at the base of the mantle or at an internal chemical interface. Plumes are initiated in the boundary layers and are sustained by thermal instabilities. The existence of a thermal boundary layer at the bottom of the mantle requires sufficient heat flow from the core. If the mantle is mainly heated from within, plumes are very weak and the associated volcanism is either insignificant or not present. In this case, mantle convection is dominated by cold downwellings and the upwellings are broad with a small temperature contrast relative to the average mantle. Mantle convection and thermal evolution models show that planetary mantles of one-plate planets are mainly dominated by internal heating during a long period of their evolution. An exception can occur in the very early evolution if the core is superheated due to core formation or if sufficient heat is generated in the core due to the decay of radioactive heat sources. For example, potassium might be incorporated in the core at high pressure (Murthy et al., 2003). In case of a superheated core, heat flow from the core decreases rapidly a few hundred million years after core formation and the temperature difference across the CMB layer decreases rapidly to only a few tens of kelvins (Breuer and Spohn, 2003, 2006; Hauck and Phillips, 2002; Spohn et al., 2001). Convection models show that plumes cease or cannot even develop under these conditions (Spohn et al., 2001). The temperature difference at the CMB, which is the maximum excess temperature of the plume, can be increased to larger values of about 150 K if the viscosity increases with depth (Buske and Christensen, 2003). The excess temperature further decreases with height above the CMB because of adiabatic cooling and heat diffusion (Albers and Christensen, 1996) and in the presence of the exothermic phase transitions, for example, α- to β-olivine and β- to γ-olivine. In conclusion, one would expect only weak volcanic activity due to present-day plumes in a planet with one-plate tectonics. However, present plume volcanism is possible for the following reasons:

1. In case of a layered mantle due to a chemical interface or an endothermic phase transition, a thermal boundary layer can form at the interface of the convecting mantle layers where upwelling plumes may develop.
2. A strong increase of viscosity with depth.
3. Heat supply from the core due to radioactive elements or gravitational energy released due to the growth of an inner core.

The maximum excess temperature of these plumes, that is, the temperature jump across the thermal boundary, however, depends strongly on the heat source distribution within the mantle and the core, the mantle rheology, and the existence of phase transitions.

In contrast to strong mantle plumes, global melting can be present for a much longer time period, probably until the present, depending on the efficiency of mantle cooling and amount of radioactive heat sources. For example, a one-plate planet with a dry mantle, that is, a mantle with a stiff rheology, a thick low-conductivity crust, and a high concentration of radioactive elements in general shows that the average mantle temperature remains close to the solidus, which is a consequence of the Tozer (1965) equilibrium described earlier in the text. A typical evolution of an associated melt zone involves the downward propagation of its upper boundary with time as a consequence of planetary cooling. Thus, the average source region of the melt migrates to greater depth.

10.08.5.1.3 Thermal properties of the crust

The crusts of terrestrial planets can act as thermal insulators for the planets. The crust is enriched in radioactive elements and its thermal conductivity is lower than that of mantle material. The thermal conductivity of most compact volcanic materials ranges between 1.5 and 3.5 $W\,m^{-1}\,K^{-1}$ at ambient temperatures (Clauser and Huenges, 1995; Clifford and Fanale, 1985), but it decreases with temperature similarly to mantle material (Seipold, 1992). For example, the thermal conductivity of typical compact basalt decreases from about 2 $W\,m^{-1}\,K^{-1}$ at 270 K to about 1.5 $W\,m^{-1}\,K^{-1}$ at 800 K. In addition to the temperature effects on thermal conductivity, the structure of the material can significantly change the thermal conductivity. Fractured and porous materials have a reduced thermal conductivity in comparison to compact material. The upper crust of the terrestrial planets and the Moon has been fractured due to impact processes, in particular in the early period of heavy bombardment. These impacts resulted in the production of a porous megaregolith that extends to considerable depth, for example, to about 10 km for Mars (Clifford, 1993) and a few tens kilometers for the Moon (Binder and Lange, 1980; Wieczorek et al., 2013). A recent analysis of high-resolution data from the GRAIL mission to the Moon has also revealed a strong lateral variation of this megaregolith layer (Wieczorek et al., 2013). At the transition between fractured and coherent basements, the lithostatic pressure is sufficient to close all fractures and the intergranular pore space. It is expected that the porosity decreases exponentially from the dusty surface to this self-compaction depth (Binder and Lange, 1980). As a consequence, thermal conductivity decreases from the bottom of the megaregolith toward the surface. At low atmospheric pressures (e.g., around the ambient atmosphere pressure on Mars and lower), surface thermal conductivity can take on extremely low values. Remote thermal measurements for the Martian surface indicate soil thermal conductivities in the range of 0.075–0.11 $W\,m^{-1}\,K^{-1}$ (Kieffer, 1976), but it is assumed that thermal conductivity increases significantly in the first few centimeters, for example, to a value of about 0.5 $W\,m^{-1}\,K^{-1}$ (Helbert and Benkhoff, 2003), typical of sedimentary material or loose sandstone. The decrease of thermal conductivity due to porosity is less if the pores are filled, for example, with water ice. Furthermore, hydrothermal circulation in a porous, permeable regolith can counteract the low conductivity of the crust. This may be relevant for Mars since there are indications that the planet had a water-rich crust for much of its geologic history.

10.08.5.1.4 Observations

To study the crustal evolution and composition of the terrestrial planets, a combination of global spectral imaging, optical imaging, gravity measurements, and laser altimetry (remote sensing), in situ investigations, and laboratory analyses of returned samples is desired. Since there is a different level of planetary exploration for each of the terrestrial planets, our knowledge varies strongly among the planets with Mars and the Moon being the most studied bodies so far.

Mapping of the surfaces using image and topography data allows us to reconstruct geologic units. Of particular importance is the determination of the relative ages of different units by crater counting. In the case of Venus, the dense atmosphere does not allow us to obtain visible images of the surface, but with the radar on board the NASA Magellan mission, it was possible to study the surface with radar images and topography data, thereby allowing crater studies. In addition, the Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) on board the Venus Express ESA mission was able to image most of the southern hemisphere in a window near 1 μm. The derived near-infrared (IR) surface emissivity provides information on differences in surface composition.

To obtain the absolute ages of the geologic units, direct samples from the planetary surfaces are required. These dated samples are available from the Moon but not for other terrestrial planets (Shergottite, Nakhlite, and Chassigny (SNC) meteorites that are believed to originate from Mars cannot be used due to the lack of knowledge of the geologic setting from where the meteorites came). Therefore, the cratering rate and the absolute ages of the Moon's surface are used to calibrate the relative surface ages of the other terrestrial planets. However, there is an uncertainty in the ratio of the specific planet/Moon cratering rates that propagates directly into uncertainty on age (**Chapter 10.10**). Despite this uncertainty, with the identification of volcanic areas and their ages, the image and topography data can also be used to estimate the amount of volcanic activity with time. However, one needs to consider that only extrusive volcanism can be observed at the surface but not intrusive material below the surface. Estimates of the relative proportion between intrusive and extrusive materials range from a factor of 7 to 10 (Greeley and Schneid, 1991), but this factor might vary strongly among the planets and the crustal formation processes. Further insight to the timing of crustal evolution can be obtained by the geochemical analysis of samples. Such studies exist for the Moon (Apollo samples) and Mars (SNC meteorites) that have

identified for both bodies an early mantle differentiation event with a major crustal formation event 4.5 Ga ago.

Seismology is the method of choice for determining crustal thickness on Earth. Seismic data are not yet available for the other terrestrial bodies (except lunar seismic data acquired during the Apollo exploration), although the deployment of seismometers on Mars is envisioned for future missions. In 2016, the NASA Discovery mission InSight plans to deploy one seismometer at the Martian surface to constrain its interior structure. In the absence of seismological data, however, the crustal thickness of terrestrial planets and the Moon is nonuniquely constrained by the gravity field and topography measurements (see **Chapter 10.05**).

The composition of the crust can be measured by a combination of different methods. The most reliable method is the examination of samples in the laboratory, possible so far for the Moon and Mars. Without direct access to samples, the mineralogical composition can also be assessed on a global scale by IR imaging spectroscopy in the range of 0.8–50 μm. This spectral domain contains a large number of signatures diagnostic of many minerals. This technique has revealed mafic to ultramafic compositions for flows on the surface of Mercury (Charlier et al., 2013) and generally mafic compositions (basalt to basaltic andesite) for the major surface units on Mars (Christensen et al., 2000; McSween et al., 2003). Mafic to ultramafic compositions are also inferred based on spectroscopically determined eruption temperatures on Io (McEwen et al., 1998).

For bodies without atmospheres, elemental compositions may be mapped from orbit using gamma or neutron spectrometry. Lunar Reconnaissance Orbiter (Lawrence et al., 1998), SELENE/Kaguya (Kobayashi et al., 2010), and Chang'E (Zhu et al., 2013) provided, for instance, maps of uranium, thorium, potassium, oxygen, magnesium, silicon, calcium, titanium, and iron. The data not only revealed correspondences with the surface features but also showed variations across previously undifferentiated units such as the far side highlands (Yamashita et al., 2010).

For in situ measurements, the alpha proton x-ray spectrometer technique has been used on the NASA Pathfinder mission to Mars and the Mars Exploration Rovers (MERs), Spirit and Opportunity, to chemically analyze soils and rocks without any sample preparation (except the use of a rock abrasion tool to analyze 'fresh' crustal material). Other in situ instruments are the Mössbauer spectrometer to determine with high accuracy the composition and abundance of iron-bearing minerals and the mini-thermal emission spectrometer to analyze the mineral composition of Martian surface features and select specific rocks and soils for more detailed investigations. Both instruments have been already used on the MER rovers. In situ measurements of the radioactivity of the Venusian crust were carried out by several Venera probes, revealing a wide variation (equivalent to granitic to basaltic compositions on Earth) in levels of heat-producing elements.

10.08.5.1.5 Incorporation of crustal growth in thermochemical evolution models

To calculate crustal growth and mantle differentiation in parameterized thermochemical evolution models, one usually assumes that both effects are regulated by the mantle convection speed (e.g., Breuer and Spohn, 2003; Fraeman and Korenaga, 2010; Grott et al., 2011a,b; Hauck and Phillips, 2002; Morschhauser et al., 2011; Schubert et al., 1990; Turcotte and Huang, 1990). In earlier models, crustal growth was simply proportional to the vigor of mantle convection and a fixed parameter measuring the efficiency of magma generation (Schubert et al., 1990; Spohn, 1991; Spohn and Breuer, 1993). These models show crustal production rates even when the mantle does not contain partial melt. In more recent parameterizations, the actual melt fraction as a function of the melting and mantle temperature is considered in more detail (Breuer and Spohn, 2003, 2006; Fraeman and Korenaga, 2010; Grott et al., 2011a,b; Hauck and Phillips, 2002; Morschhauser et al., 2011). The parameterizations in the literature are slightly different, for example, in how they incorporate the increase of the melting temperature due to continuous depletion of the mantle in crust-forming elements and how they include volcanic heat transport through the lithosphere (see succeeding text). However, at least in the case of Mars, they basically give similar results (Breuer and Spohn, 2003, 2006; Hauck and Phillips, 2002).

In the early evolution of a planet, typically the heat flow out of the lithosphere that is transported from the interior to the surface via volcanism is larger than the heat transported conductively. This mode of planetary heat transport is called the heat-pipe mode (O'Reilly and Davies, 1981) after the localized channels through which melt brings heat to the surface. Heat pipes are conduits that transfer heat and material from the base of the lithosphere to the surface via buoyant ascent. The lithosphere is then advected downward due to the volcanic loading and cools more efficiently, that is, the lithosphere is thickened and the geotherm is depressed by the advection. This cooling mechanism plays a substantial role even today for Jupiter's moon Io (Moore, 2001; O'Reilly and Davies, 1981) and in the early evolution for the terrestrial planets (Moore and Webb, 2013; Morschhauser et al., 2011; Spohn, 1991). Recently, it has been shown with 2-D convection models that in addition to altering the thickness and thermal structure of the lithosphere, heat pipes also influence the stress distribution (Moore and Webb, 2013). Lithospheric stresses are reduced by the combination of thickening and the loss of buoyancy from the convective thermal boundary layer at the top of the mantle through the heat pipes. As a consequence, the initiation of plate tectonics in the case of the Earth was possibly triggered at the end of the heat-pipe mode about 3.2 Ga ago due to increased lithospheric stresses sufficient to exceed the yield stress of the plate (Moore and Webb, 2013).

More recently, the thermochemical evolution of terrestrial planets, in particular for Mars, has been also calculated with 2-D and 3-D convection models (e.g., Golabek et al., 2011; Keller and Tackley, 2009; Nakagawa and Tackley, 2012; Ogawa and Yanagisawa, 2011; Plesa and Breuer, 2014; Ruedas et al., 2013a,b). These models do not require parameterizing the crustal growth with the convective vigor but instead calculating self-consistently the melt production and associated processes like depletion of heat sources and volatiles in the residual mantle whenever the local temperature exceeds the solidus temperature. In most models, it is assumed for simplicity that partial melt is instantaneously removed from the interior, although a few global models also consider the transport of melt via two-phase flow (Ogawa and Yanagisawa, 2011), and most models so far neglect the additional cooling due the downward migration of the lithosphere that can be important

for high crustal production rates (Moore and Webb, 2013). The models allow consideration of regional effects of melting that are necessary to study, for instance, the crustal dichotomy and large volcanic provinces of Mars and the mare volcanism on the Moon. Nevertheless, general trends as with the parameterized convection models have been identified with the 2-D and 3-D convection models, for instance, the interior temperature of a planet is significantly buffered in the case of melting due to the consumption of latent heat, the efficient transport of melt toward the surface, and the depletion of radioactive elements in the mantle.

10.08.5.2 Magnetic Field Generation

Magnetic field generation in the iron-rich core of a planet is strongly coupled to the thermal evolution of the planet. Thus, knowledge about the magnetic field evolution can be used to constrain the thermal evolution of a planet. Planetary space missions instrumented with magnetometers provide insight about the present magnetic fields of the terrestrial planets. The available data, however, differ significantly. In the case of the Moon, the lunar magnetic field has been studied indirectly via the natural remanent magnetization of the returned lunar samples and directly with magnetometers carried to the surface and placed in orbit at low altitude above the surface on the Apollo 15 and Apollo 16 subsatellites (see reviews in Dyal et al., 1974; Fuller and Cisowski, 1987). In 1998, Lunar Prospector allowed detailed global measurements of the magnetic field (Lin et al., 1998). In the case of Mars, Mars Global Surveyor (MGS) measured the magnetic field of the entire planet (Acuña et al., 1998). Both bodies, that is, the Moon and Mars, show no sign of a present internal field but parts of their crusts preserve a remanent magnetization (Acuña et al., 1999; Coleman and Russell, 1977). It is assumed that the remanent magnetization was established at the time of crustal formation, while an internal dynamo was active (Acuña et al., 1999; Runcorn, 1975). An age determination of the magnetized and nonmagnetized crust would thus allow us to reconstruct the magnetic field history. Estimates have been made for the Moon (e.g., Garrick-Bethell et al., 2009; Shea et al., 2012; Tikoo et al., 2012) and Mars (e.g., Arkani-Hamed and Boutin, 2012; Lillis et al., 2008, Mitchell et al., 2007).

In the case of Mercury, the two flybys by Mariner 10 have identified a weak internal field (Ness et al., 1974) – an observation recently confirmed by the NASA MESSENGER mission (Anderson et al., 2011). Venus, like Mars and the Moon, has no present-day dynamo action in its core. For both planets, Mercury and Venus, we cannot exclude the presence of an old magnetized crust because data are not available so far, but the temperature of the crust of Venus is likely close to the Curie point of magnetic minerals even at the surface suggesting – if existent at all – a thin magnetized crustal layer. As long as we do not know whether the crust is magnetized, we are unable to reconstruct the magnetic field history.

10.08.5.2.1 Core convection and necessary conditions for dynamos

To generate a self-sustained magnetic field, convection within the electrically conductive, metallic core is required. To obtain core convection, one must appeal to unusually large heat flows or the development of an inner core. In either case, the core must be cooling and the cooling is controlled by the heat transport of the outer layers, that is, the mantle and the crust (for a review, see Breuer et al., 2010).

10.08.5.2.1.1 Thermal dynamo

Thermal convection in the core, like thermal convection in the mantle, is driven by a sufficiently large superadiabatic temperature difference between the core and the mantle. It occurs if the core heat flow exceeds that conducted along the core adiabat. The latter heat flow, therefore, serves as a criterion for the existence of thermally driven convection in the core:

$$q_{\mathrm{crit}} = k_c \frac{\mathrm{d}T}{\mathrm{d}r} = k_c \frac{\mathrm{d}T}{\mathrm{d}P}\frac{\mathrm{d}P}{\mathrm{d}r} = k_c \frac{\alpha_c g_c T}{C_c} \qquad [22]$$

with k_c the thermal conductivity of the core, $\mathrm{d}T/\mathrm{d}r$ the temperature gradient in the core, P the pressure, α_c the thermal expansivity, C_c the heat capacity, and g_c the gravity. The critical heat flow varies between 5 and 20 mW m^{-2} for Mars (Nimmo and Stevenson, 2000) and Mercury and between 15 and 40 mW m^{-2} for Venus and is smaller than 7 mW m^{-2} for the Moon. The large variation in the value for each planet comes partly from the uncertainty in the thermal conductivity of the core material, which varies between 43 and 88 W m^{-1} K^{-1} at the core pressures (Anderson, 1998; Stacey and Anderson, 2001). However, it should be noted that recent advances in high-pressure mineral physics have indicated that the thermal conductivity (and thus the minimum heat flux for convection) of the Earth's core is likely a factor of 2 higher than previously thought – and thus higher values are also possible for other terrestrial planets. Calculated values of the thermal conductivity of the Earth's core range from 90 to 150 W m^{-1} K^{-1} (de Koker et al., 2012; Gomi et al., 2013; Pozzo et al., 2012).

The cooling of the core is strongly controlled by the thermal evolution of the overlying mantle. If the mantle removes heat from the core at a rate that exceeds the critical heat flow, then the core will convect. If the mantle removes heat at a rate less than the critical heat flow, the core is thermally stably stratified and dynamo action by thermal convection is not possible.

10.08.5.2.1.2 Chemical dynamo

Compositional convection can occur due to the release of positively buoyant material during the process of solid inner core freezing from a fluid core with noneutectic composition (Braginsky, 1964). The existence of light elements in the core has been assumed for all terrestrial planets (**Chapter 10.02**), and many suggestions have been offered for the amount and identity of the light elements that are mixed with the iron. For example, sulfur in the core – just as most other light elements in small enough concentrations – has two major effects: first, a reduction of the average density of the core and, second, a significant reduction of the melting temperature (e.g., Fei et al., 1997; Usselman, 1975). For most terrestrial planets, it is usually assumed that the composition of the sulfur-bearing core is on the Fe-rich side of the eutectic composition (e.g., Lodders and Fegley, 1998), whereas for the large Jovian moons (e.g., Scott et al., 2002) and Mars (e.g., Stewart et al., 2007), this is debated.

The conventional model of inner core growth on the Fe-rich side of the eutectic is applicable in cases where the (positive) slope of the core liquidus is steeper than that of the adiabat

(e.g., Stevenson et al., 1983). As a consequence, when the core has cooled to reach the liquidus at the center, iron will crystallize and an inner core nucleus will form and – upon further cooling – the core will freeze from the inside out. The outer core will be increasingly depleted in iron. The concentration of sulfur will not be homogeneous, though. Rather, a boundary layer will form just above the inner core where the sulfur concentration is above average. The fluid in this layer will be lighter than the remainder of the fluid core and compositional convection to homogenize the composition will be induced. Chemical convection can drive a dynamo more effectively than thermal convection (Braginsky, 1964; Stevenson et al., 1983). In contrast to the transformation of thermal energy to magnetic field energy, where only a small part of the energy is transferred (described by the Carnot efficiency factor, which is between 0.06 and 0.11; Lister and Buffett, 1995; Stevenson et al., 1983), a chemically driven dynamo has no similarly reduced efficiency.

Chemical convection and the associated generation of a magnetic field in the core occur if the temperature in the fluid (outer) core lies between the solidus and the liquidus of the core material. Inner core growth permits outer core convection even when the heat flow through the CMB is less than the heat carried by conduction along the adiabat.

Compositional convection, however, can be very different for sulfur-rich compositions or in smaller planetary bodies. For a sulfur-rich composition, solid FeS (or Fe_3S for pressures higher than 20 GPa) might precipitate deep within the core (Hauck et al., 2006; Stewart et al., 2007). Because of the high sulfur content of FeS (>36.5 wt% S), it is less dense than the residual liquid. The solid FeS thus buoyantly rises and remelts at higher levels. With decreasing temperature, the previous floatation and remelting of FeS increase the sulfur content near the CMB enough to raise the melting point of the liquid such that solid FeS is stable near the CMB resulting in the process of precipitation of FeS both deep and near the CMB (Hauck et al., 2006). The scenario described results in dynamics that are different from the Fe-rich side. In case of solid Fe_3S, it is assumed to be denser than the residual liquid and remains in the center to form an inner core as in the conventional Earth scenario (Stewart et al., 2007).

Recent experimental works (e.g., Chen et al., 2008; Chudinovskikh and Boehler, 2007; Fei et al., 1997, 2000; Stewart et al., 2007) have further identified two important aspects of the Fe–FeS phase diagram: (1) At pressures less than 14 GPa, the eutectic melting temperature decreases with increasing pressure, and (2) the S content at the eutectic decreases with increasing pressure.

A negative slope melting temperature in the core will have profound implication. For instance, Fe may precipitate at the CMB rather than in the center and may fall as iron snow. The solid iron particles are heavier than the surrounding Fe–FeS fluid, that is, a snow zone forms; settle to deeper core regions, where the core temperature is higher than the melting temperature; and remelt again. As a consequence, a stable chemical gradient in the Fe–FeS fluid arises within the snow zone (Hauck et al., 2006; Williams, 2009). The thickness of the snow zone increases with time, and below that layer, the fluid inner core is continuously enriched with remelting iron. This so-called iron snow regime process has been already suggested for the present magnetic field in Ganymede (Hauck et al., 2006; Rückriemen et al., 2014) and Mercury (Chen et al., 2008; Vilim et al., 2010), for a possible future chemical dynamo in Mars (Stewart et al., 2007) or for the early magnetic field on the Moon (Zhang et al., 2013) – although it is debated how and where a magnetic field can be generated in such a regime. Zhan and Schubert (2012) modeled fluid motions within the snow zone assuming that Fe snow resembles internally heated convection and obtain multipolar solutions for the magnetic field. On the other hand, Bland et al. (2009) and Williams (2009) argued that compositional-driven convection in the Fe snow zone does not provide enough energy to generate a magnetic field, while Rückriemen et al. (2014) argued that the style of convection via sedimentation is too small-scale and therefore lacks an important criterion necessary for dynamo action (Christensen and Wicht, 2007) as also shown by the dynamo models of Vilim et al. (2010). However, below the Fe snow zone, the process of remelting of iron creates a gravitationally unstable situation and it is proposed that this could be the driving mechanism for a potential dynamo (Rückriemen et al., 2014; Vilim et al., 2010). This dynamo scenario would be restricted to the time period the snow zone needs to grow across the core. Note that in the case of Mercury, it has been even suggested that multiple snow layers can exist in the core (Chen et al., 2008; Vilim et al., 2010).

In conclusion, whether a terrestrial planet has core convection and whether it can have a thermal or chemical dynamo depends on the cooling rate of the core, the core temperatures, and the core melting temperatures. If the mantle transports sufficient heat, then a dynamo occurs, but if the mantle is too insulating, then there is no dynamo.

Thermal evolution models indicate that the existence of a present-day magnetic field for a terrestrial planet is associated with the growth of an inner core. A present-day thermal dynamo is unlikely since the results show a slow cooling of a core with heat flow less than the heat flow along the core adiabat. In fact, a thermal dynamo is in general difficult to generate and only likely if the core is superheated with respect to the mantle due to the core formation process. If a thermal dynamo exists, it shuts off very early in the evolution as the heat flow out of the core decreases rapidly during the first few hundred million years. Whether in the subsequent evolution an inner core can grow and thus a compositional dynamo can be initiated depends strongly on the temperature evolution and the core composition.

10.08.6 Comparison of the Terrestrial Planets and the Moon

Although the evolution, composition, and interiors of terrestrial planets have some features in common, such as iron-rich cores, slowly convecting silicate mantles, and rigid crustal layers – they are entirely different in many other respects.

For instance, a strong diversity in crustal evolution has been identified: Mercury and the Moon show highly cratered surfaces, indicating an average crust older than about 4 Ga with only minor volcanic activity in the subsequent evolution. Mars shows a distinct dichotomy of the surface consisting of the old southern highlands with an average age of more than

4 Ga and the superficially younger northern lowlands with an average age of about 3.5 Ga (Tanaka et al., 1992). Below the comparative young surface, the northern crust has a similar age as the southern crust as concluded from topography data of partly buried impact structures (Frey et al., 2002). A continuous and rapid decline of global volcanism and crustal growth since planet formation can be observed in the surface morphology. Over time, global volcanism has been concentrated into mainly two volcanic regions, Tharsis and Elysium. In comparison to the other terrestrial planets, except the Earth, the average surface of Venus seems to be the youngest with an average crustal age of 300–800 Ma (McKinnon et al., 1997; Schaber et al., 1992).

Similar to crustal evolution, magnetic field evolution differs among the terrestrial planets and the Moon. Mercury is the only terrestrial planet, except the Earth, that has a magnetic field (Ness et al., 1974) – although a weak one – that is generated in the iron core (Jupiter's satellite Ganymede also shows an internally generated field; Kivelson et al., 1996). Venus, Mars, and the Moon on the other hand have no present global magnetic fields, but at least in the case of the Moon and Mars, the remanent magnetization of part of their crust indicates that they once had dynamo action in their cores (Acuña et al., 1999; Runcorn, 1975). The diversity in the volcanic and magnetic history of the bodies is also reflected in their thermochemical evolution. In the following sections, we will describe the thermal, volcanic, and magnetic evolution of the planets.

10.08.6.1 Mercury

After the three flybys of Mercury by the Mariner 10 spacecraft in 1974–75 that returned images of about 45% of the planet's surface, a wealth of information has been obtained since 2011 by the polar orbiter MESSENGER – information that not only confirmed our view of Mercury but also revised some previous assumptions about the interior structure, dynamics, and evolution of the interior.

The images of both missions revealed a lunar-like terrain with an average age of more than 4 Ga and lobate scarps that are more or less evenly distributed over the surface. Initially, the scarps were thought to indicate an average contraction of the planet's radius by 1–2 km since the end of heavy bombardment (Strom et al., 1975). Later, estimates of the contractional strains associated with lobate scarps as imaged both by Mariner 10 and by MESSENGER during its flybys implied a limited decrease of planetary radius of up to 0.8 km only (Watters and Nimmo, 2010; Watters et al., 1998, 2009). However, the data can be strongly biased by the illumination conditions at which the images were taken (Watters et al., 2009) and recent data collected during MESSENGER's orbital phase at high Sun incidence angles allowed Di Achille et al. (2012) to identify a number of previously undetected structures. These imply a revised value of the overall radius decrease of 2.8–3.6 km if wrinkle ridges are also taken into account, and in the most recent comprehensive study of contractional tectonics, an even larger value of 5–7 km has been inferred (Byrne et al., 2014).

The overall radius decrease provides an important constraint on the planet's thermal evolution. The most likely source of global contraction is a combination of a thermal contraction (reduction in average internal temperature) and a phase change by solidification (e.g., growth of an inner core). Core freezing contributes more to the global contraction. Complete core solidification would shrink the planet about 17 km in radius (Solomon, 1976), much more than required from the geologic observations. This effect can be considerably amplified if a light alloying element such as sulfur is present in the core. Light elements will be expelled from the inner core upon solidification, increasing the density difference between the liquid and solid core phases. A small inner core is therefore better to reconcile with a small planetary contraction. The total planetary radius change as a function of inner core size for different core bulk sulfur contents is shown in **Figure 5**. Already small inner core sizes of $R_i/R_c = 0.2$ result in global contraction of 1 km. For $R_i/R_c = 0.4$, the associated global contraction already exceeds the amount of contraction observed on the surface, posing severe constraints on the admissible inner core sizes.

Thermal evolution models (Grott et al., 2011a; Hauck et al., 2004; King, 2008; Schubert et al., 1988; Tosi et al., 2013a) indicate that the observed scarps strongly restrict the possible evolution scenarios of Mercury. One possible explanation is that much of the contraction would have predated the observable geologic record, for instance, most of the inner core was formed before the end of heavy bombardment (Schubert et al., 1988). Alternatively, the inner core started to grow only recently and is small (Grott et al., 2011a; Hauck et al., 2004; Tosi et al., 2013a). In any case, to meet the constraint of only 2.4–3.6 km of contraction since the end of heavy bombardment, relatively little cooling of the interior should have happened since that time, which requires a relatively stiff mantle (high reference viscosity), a low enrichment of radioactive elements in the crust, and a thin crust (Grott et al., 2011a; Tosi et al., 2013a). With the more relaxed constraint allowing a global contraction between 5–7 km (Byrne et al., 2014), these requirements are still valid but a slightly larger parameter variation is possible than shown in **Figure 6**. In particular, a more volatile-rich mantle, thus a lower reference viscosity is possible (Tosi, personal communication).

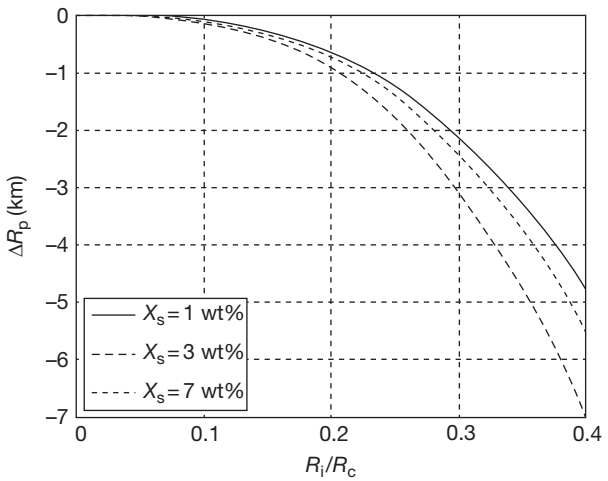


Figure 5 Total radius change ΔR_p as a function of inner core size R_i normalized to the core radius R_c for different initial core sulfur contents χ_S. Adapted from Grott M, Breuer D, and Laneuville M (2011a) Thermochemical evolution and global contraction of Mercury. *Earth and Planetary Science Letters* 307: 135–146. http://dx.doi.org/10.1016/j.epsl.2011.04.040.

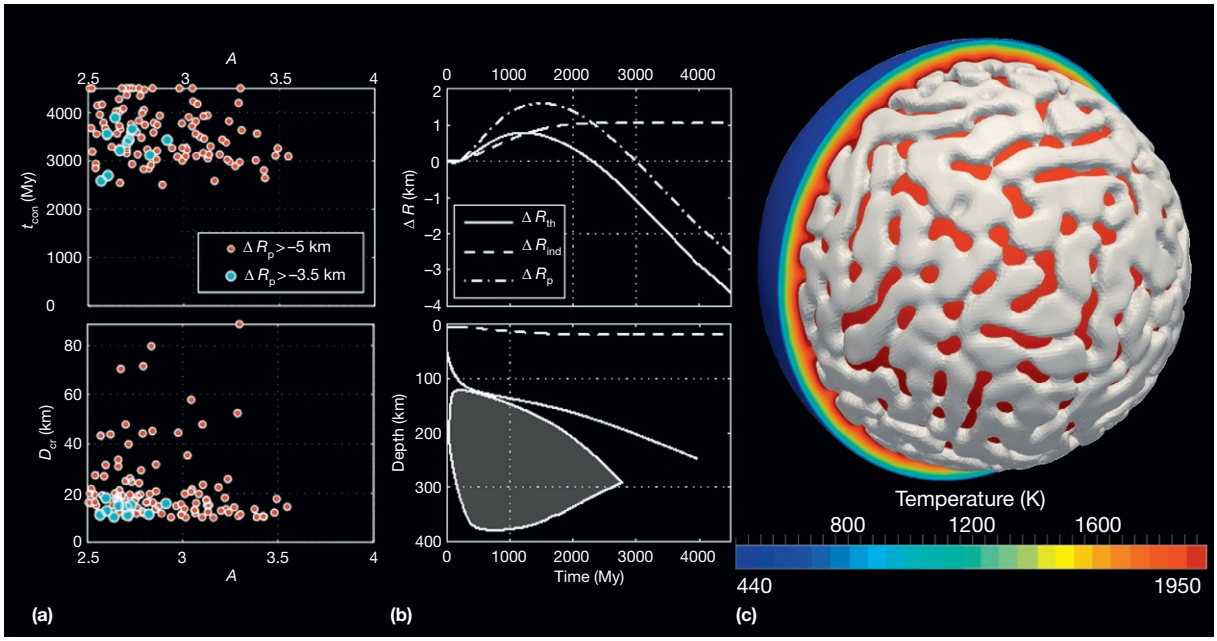


Figure 6 (a) Monte Carlo simulations: (top) convection duration and (bottom) present-day crustal thickness depending on the enrichment of radiogenic heat sources in the crust. (b) Typical thermal evolution: (top) different contributions ΔR (contribution of thermal expansion and contraction ΔR_{th} and of mantle differentiation due to partial melting ΔR_{md}) to the global planetary radius change ΔR_p as a function of time; (bottom) crustal thickness, stagnant lid thickness, and partial melt zone (gray area) as function of time. (c) 1900 K temperature isosurface of a three-dimensional simulation. Tosi N, Grott M, Plesa A-C, and Breuer D (2013) Thermo-chemical evolution of Mercury's interior. *Journal of Geophysical Research*: *Planets* 118: 2474–2487. http://dx.doi.org/10.1002/jgre.20168.

The thermal history of terrestrial planets depends strongly on the amount of radioactive elements in their interiors. Predictions for the abundances of the important heat-producing elements uranium, thorium, and potassium in Mercury depend on the formation scenario, which seems to be very special among the terrestrial planet, as indicated by Mercury's large bulk density of ~5430 kg m^{-3}. There are three general hypotheses for the formation of Mercury. The first scenario for forming Mercury's large bulk density involves aerodynamic sorting of iron and silicate particles in the solar nebula (Weidenschilling, 1978) that might have given the planet abundances of U and Th similar to the upper mantle of the Earth, though lacking much K (Basaltic Volcanism Study Project, 1981). The other two models for Mercury's formation involve a later stage loss of silicate, and retention of iron, materials. Impact-induced stripping of the outer, silicate layers of a larger, differentiated, proto-Mercury (e.g., Cameron et al., 1988; Wetherill, 1988) could have left a silicate layer with near CI chondritic abundances of heat-producing elements (e.g., Lodders and Fegley, 1998). A final, more exotic scenario involves a late-state vaporization of silicates that would result in a Th-rich silicate layer lacking appreciable U and K (Fegley and Cameron, 1987). Of the three formation models, the vaporization model predicts less cooling and contraction of the planet (Hauck et al., 2004) as it assumes ^{232}Th as the sole of radiogenic heat (Fegley and Cameron, 1987), which with its 14 Gy half-life has not had a significant decline in heat output. However, evidences for pyroclastic eruptions are documented in MESSENGER images and high magma volatile contents above >0.36 wt% are required to emplace pyroclasts to the observed distances (Kerber et al., 2011). A more volatile-rich content as previously assumed is also suggested by the measurements of the average surface abundance of the radioactive isotopes by the MESSENGER gamma-ray spectrometer (Peplowski et al., 2011, 2012). Both hint at a volatile content similar to other terrestrial planets and hence permit to rule out at least vaporization and condensation models as these require a nearly complete loss of volatiles.

To address this problem, Grott et al., (2011a) and Tosi et al. (2013a) reinvestigated the thermochemical evolution of Mercury using a nonrefractory compositional model, taking the presence of a thermally insulating regolith layer into account. Models with a stiff mantle rheology satisfy the observational constraints of contraction if the regolith layer is at least 2 km thick. Inefficient mantle convection and thermal insulation significantly slow planetary cooling and prolong the phase of crustal production to 2.5 Gy after core formation. The volume increase associated with mantle differentiation then offsets some of the radial contraction caused by planetary cooling (**Figure 6**). It should be noted though, at the relatively low oxygen fugacities of Mercury's interior, that can be inferred from the high S and low FeO content of the surface (Zolotov et al., 2013), elements such as K, Th, and U, which are usually lithophile, can become more siderophile (Malavergne et al., 2010). McCubbin et al. (2012a) thus pointed out that it is still uncertain whether Mercury's mantle is enriched in volatiles or not.

Low-degree coefficients of the gravity field obtained from the MESSENGER radio science experiment (Smith et al., 2012) combined with Earth-based radar measurements of the planet's spin state (Margot et al., 2007) have been used to revise the internal mass distribution (see also **Chapter 10.02**).

Before MESSENGER, Mercury's high density (e.g., Anderson et al., 1987) indicated already a massive core (e.g., Siegfried and Solomon, 1974). While a core radius of about 1800–1860 km (consistent to a mantle thickness of about 580–640 km) has previously been assumed in most thermal evolution models (e.g., Grott et al., 2011a; Hauck et al., 2004; Schubert et al., 1988), the new data and current interior structure models predict a significantly larger core (Hauck et al., 2013; Riner et al., 2008; Rivoldini et al., 2009). Mantle thicknesses between 300 and 500 km are possible, with intermediate values around 400 km being the most likely.

In general, a thinner mantle cools more efficiently than a thicker mantle and convection in a thin mantle possibly ceases faster. The need to understand whether mantle convection is currently ongoing or ceased in the past is also related to the interpretation of the gravity field and topography in terms of the interior structure of the planet. If the mantle is still convecting, deep-seated density anomalies and flow-induced dynamic topography of the surface and CMB may contribute to the observed fields. Michel et al. (2013) recently presented a parametric study of 2-D convection models. They analyzed the effects of different core radii between 1800 and 2100 km on the longevity of mantle convection and concluded that, for a sufficiently high rate of internal heat generation, convection can persist until the present day in a mantle at least 400 km thick. However, to further fulfill the constraint of small planetary contraction for values up to 5 km, Tosi et al. (2013a) argued that with a thin mantle of 400 km, present-day convection is unlikely. For the revised contraction value of 5–7km (Byrne et al., 2014), convection is still unlikely for a dry mantle but possible for a volatile-rich mantle (Tosi, personal communication). The planform of mantle convection in a thin mantle typically consists of several small-scale cells (**Figure 6**; Breuer et al., 2007; Tosi et al., 2013a). Regular long-wavelength patterns of linear rolls that were held responsible for the orientation of contractional strain associated with lobate scarp formation (King, 2008) can only be obtained with specific initial low-degree perturbations of the temperature field (Tosi et al., 2013a).

Due to the specific interior structure of Mercury, that is, the relatively thin mantle and the associated efficient cooling, it has been speculated that a pure iron core could be totally frozen very early in the evolution (Siegfried and Solomon, 1974; Solomon, 1976). Such a scenario is consistent with the geologic observation of small global contraction since the end of the heavy bombardment. A totally frozen core, however, is difficult to reconcile with the magnetic field history of Mercury. During the first and third close encounters in 1974 and 1975, Mariner 10 passed briefly through a small but earthlike magnetosphere. The analyses of these observations revealed the presence of an internal field, with a dipole moment that is a factor of about 10^4 smaller than that of the Earth (Ness et al., 1974, 1975, 1976; Russell et al., 1988; see review in Connerney and Ness, 1988) – findings that have been confirmed with the MESSENGER mission (Anderson et al., 2011). The discovery of the internal field strongly suggests the presence of a fluid outer core. To prevent the core from freezing, the addition of radioactive heat sources into the core (Toksöz et al., 1978), a late core formation (Sharpe and Strangway, 1976; Solomon, 1977), or deep mantle heat sources (Cassen et al., 1976) has been suggested. Late core formation would support a cool initial state for Mercury but is at variance with accretion models (e.g., Schubert et al., 1988). The most likely reason for Mercury not having a totally frozen core, however, is the incorporation of a light alloying element into its core that reduces the core melting temperature. Here, sulfur is the most likely candidate (McCammon et al., 1983; Ringwood, 1977) although strong chemically reducing conditions indicated by MESSENGER XRS measurements of surface materials (Nittler et al., 2011) suggest a core containing Si and S (Malavergne et al., 2010) that may have strong implication on the core structure and evolution (Hauck et al., 2013; Smith et al., 2012).

The amount of sulfur depends on the formation scenario of Mercury. In the most conservative equilibrium condensation models, there is no sulfur at all in the Mercurian core due to the close position of the planet to the Sun (Grossman, 1972; Lewis, 1972), an assumption mostly used in earlier thermal evolution models (Sharpe and Strangway, 1976; Siegfried and Solomon, 1974; Solomon, 1976, 1977). However, a low concentration of sulfur has been suggested with radial mixing of planetesimals and/or nonequilibrium condensation models (Basaltic Volcanism Study Project, 1981; Wetherill, 1985). Mercury's inner structure that is consistent with its high density can actually be explained with a wide range of the sulfur concentration in the core ranging from a pure iron core to a core with eutectic Fe–FeS composition (Harder and Schubert, 2001).

The observation of the magnetic field provides probably the best constraint on the interior evolution, and assuming that it is indeed generated in the core and not a remnant field – as on Mars – it proves that there is a present convecting outer core. The existence of the fluid outer core places constraints on the maximal sulfur content in the core. The inferred small planetary contraction from the lobate scarp, on the other hand, places constraints on the minimal amount of sulfur in the core (Breuer et al., 2007; Grott et al., 2011a; Hauck et al., 2004; Tosi et al., 2013a). In general, one can state that the less efficient the planetary cooling, the smaller is the required sulfur content.

Thermochemical evolution models based on temperature-dependent viscosity laws that couple the thermal, magmatic, and tectonic evolution of Mercury (Hauck et al., 2004) indicate that a dry non-Newtonian mantle and early crustal formation allows the planet to cool more efficiently than a dry Newtonian mantle. The more crust is produced, the stronger is the mantle cooling and the associated depletion of the mantle in heat sources. As a consequence, acceptable models for the sulfur content range between 6.5 and 7.5 wt% S to explain the observations and require a late formation of the inner core. A similar result has been suggested by the thermochemical models of Grott et al. (2011a) and Tosi et al. (2013a), further considering the isolating effect of the regolith layer and a nonrefractory composition.

The thermal evolution models (Buske, 2006; Grott et al., 2011a; Hauck et al., 2004; Schubert et al., 1988; Tosi et al., 2013a) show a typical evolution of the magnetic field: in the early evolution, a dynamo is generated by thermal convection before the start of the inner core and a chemical dynamo. Whether the dynamo action ceases for a time until inner core growth provides sufficient energy depends on the onset time of inner core growth. The higher the sulfur content, the later is the onset of inner core growth (e.g., Grott et al., 2011a; Hauck et al., 2004; Schubert et al., 1988; Tosi et al., 2013a). A detection of remanent magnetized crust and the

determination of its age by future missions could therefore help to constrain the amount of sulfur in the core and thus the formation scenario.

The hydromagnetic dynamo as assumed in the thermal evolution models in the preceding text poses a problem in the strength of the calculated field. Although there is presently no satisfying parameterization to calculate the magnetic field strength, estimates suggest a much larger field than that observed (Schubert et al., 1988; Stevenson et al., 1983). Recent thin-shell dynamo models, however, have shown that planets with a large solid inner core (relative inner core size larger than about 0.8) can produce magnetic fields with Mercury-like field intensities (Heimpel et al., 2005; Stanley et al., 2005). Alternatively, a deep dynamo model that suggests a rather small inner core (smaller than 1000 km in radius) has been proposed by Christensen (2006). In this model, the dynamo operates only at depth and the associated dynamo field is strongly attenuated by the skin effect through a stable conducting region of the upper core but can only explain the observed strength of the surface magnetic field when the sulfur concentration in Mercury's core is only a fraction of a percent (Manglik et al., 2010) – possibly not consistent with the thermochemical evolution of Mercury (Grott et al., 2011a; Hauck et al., 2004; Tosi et al., 2013a).

All the latter mentioned dynamo models assume the common earthlike dynamo for which iron starts to crystallize in the core center. The melting behavior of the iron–sulfur system at moderate pressures, however, suggest a negative liquidus temperature gradient under conditions found at shallow depths in Mercury's core. At the present time, the core is most likely precipitating solid iron in the form of snow, at a single depth or in two distinct zones (Chen et al., 2008). Dynamo models considering an iron snow zone midway through the core can produce the observed field strength (Vilim et al., 2010). Considering a Fe–S–Si ternary system in the core (e.g., Malavergne et al., 2010), experiments have demonstrated the immiscibility of S- and Si-rich liquids up to a pressure of ~15 GPa (Morard and Katsura, 2010; Sanloup and Fei, 2004) and smaller density differences between liquid and solid compared to the Fe–S binary system (Kuwayama and Hirose, 2004). These findings suggest a compositional stratification in the core as S-rich liquids would segregate toward the CMB leaving more Si-rich materials deeper (Hauck et al., 2013; Smith et al., 2012). The consequences of such a system on the dynamo generation are, however, unknown.

Another alternative way of generating a weak magnetic field in Mercury is by a thermoelectric dynamo (Giampieri and Balogh, 2002; Stevenson, 1987). This dynamo makes use of a thermoelectric electromotive force set up at a distorted CMB. Such a dynamo requires topography variations of the CMB of the order of a few kilometers due to mantle convection. Which kind of dynamo is active in Mercury cannot be concluded from the current magnetic data. New insights into the magnetic field evolution and the dynamo mechanism of Mercury are expected from the coming BepiColombo mission.

Among the new findings by MESSENGER is the observation that the volcanism probably lasted well into the second half of solar system history (Head et al., 2011; Prockter et al., 2010), while Mariner 10 images showed no evidence for substantial additions to the crust after the end of the late heavy bombardment (e.g., Spudis and Guest, 1988). Volcanic activity that extends a few billion years can be obtained considering a thermally isolating crust and regolith layer (Breuer et al., 2007; Grott et al., 2011a). Furthermore, x-ray spectrometric measurements of the surface materials revealed higher Mg/Si and lower Al/Si and Ca/Si ratios than typical terrestrial or lunar crusts (Nittler et al., 2011). These ratios are indicative of compositions intermediate between basaltic and ultramafic rock types, possibly comparable to terrestrial komatiites, and thus hinting at an origin from partial melts at melt fractions as high as 20–30%. Estimates of the crustal thickness have a large possible range. Anderson et al. (1996) used the observed center-of-mass and center-of-figure offset together with an assumption of Airy isostasy to infer a crustal thickness of 100–300 km. Further, constraints on the crustal thickness have been derived from viscous topographic relaxation (Nimmo, 2002). If the faulting observations provide a reasonable guide to heat fluxes on Mercury at 4 Ga, topography is most likely to have survived if the mean crustal thickness were less than 200 km. Studies employing topographical information about the Mercurian surface (Nimmo and Watters, 2004; Watters et al., 2004) suggest that the extensional faulting of observed lobate scarps is consistent with a mean crustal thickness of 90–140 km. The most recent gravity and topography data from MESSENGER by Smith et al. (2012) suggest an average value of 50 km assuming a density contrast of only 200 kg m^{-3} between the mantle and crust.

Thermochemical evolution models with parameterized stagnant lid convection and assuming an earthlike mantle for Mercury derive crustal thicknesses that vary between 5 and 70 km depending on the initial temperature distribution, the assumed amount of radioactive elements, and the mantle rheology (Breuer et al., 2007; Grott et al., 2011a; Hauck et al., 2004; Tosi et al., 2013a). However, with a mantle thickness of ~400 km, crustal thicknesses between 10 and 20 km are more likely (**Figure 6**) as a thicker crust tends to result in stronger contraction after the LHB, inconsistent with the observations.

10.08.6.2 Venus

The statistically uniform distribution of impact craters prevents us from identifying older or younger regions and thus from determining resurfacing rates. The surface has a global average age of 300–800 Ma (McKinnon et al., 1997; Schaber et al., 1992). From geochemical measurements of the Soviet landers, the surface appears to be basaltic (McKenzie et al., 1992) although there is a range of compositions, an observation that is supported by Magellan radar images. These images show many long volcanic flows that are similar to basaltic flows on Earth. It has been suggested that the entire surface was covered by a volcanic event of global-scale (McKinnon et al., 1997), but such an event is not required by the crater distribution (Bjonnes et al., 2012). In such models, the resurfacing event was followed by almost no volcanic activity to the present day. The cause for this global resurfacing event and whether it was a single event or one of a periodically recurring set of events is highly uncertain. Thus, a number of papers that speculate on the global resurfacing event can be found in the literature (e.g., Bougher et al., 1997; Nimmo and McKenzie, 1998; Turcotte, 1993):

1. Widespread melting is caused by a transition from layered to whole mantle convection (Steinbach and Yuen, 1992). As Venus is cooling, the tendency for mantle layering that is caused by the endothermic spinel-to-perovskite phase transition decreases until the layered structure breaks up with a single catastrophic event.
2. Periodic instabilities of a depleted layer beneath the crust (Herrick and Parmentier, 1994; Parmentier and Hess, 1992) result in the global melting event. In that scenario, the competition between compositional stratification due to crustal formation and thermal buoyancy may cause episodic overturn of the mantle.
3. The resurfacing is associated with a transition from oscillatory to steady convection due to planetary cooling and thus a decreasing Rayleigh number (Arkani-Hamed et al., 1993).
4. The global resurfacing has been related to some kind of episodic plate tectonics. After a resurfacing event, the cold lithosphere has been subducted, and a new lid grows conductively. As this lid thickens, its negative buoyancy increases, until at some critical thickness, it is able to subduct again and thereby instigates a new resurfacing event (Armann and Tackley, 2012; Fowler and O'Brien, 1996; Moresi and Solomatov, 1998; Turcotte, 1993; Weinstein, 1996). Herrick (1994) concluded that a change in surface boundary condition could lead to a rapid rise in mantle temperature and catastrophic resurfacing in the form of plate subduction.
5. Solomatov and Moresi (1996) assumed that the resurfacing event is caused by the cessation of plate tectonics. The reduction in convective stresses with time might cause deformation in the stagnant lid to change from brittle (plate tectonics) to ductile (little surface movement).

The commonly assumed rapid global resurfacing event has been questioned by Campbell et al. (1999) and more recently by Bond and Warner (2006) and Bjonnes et al. (2012). According to the latter authors, the cratering record allows a variety of interpretations in terms of volcanic resurfacing including a global decrease with time in the rate of an otherwise statistically distributed volcanic activity. Previous authors had concluded that Venus underwent a major transition in tectonic style albeit more gradual than the previously postulated sudden global resurfacing event. The surface geology also seems to indicate that modifications of the surface are planetwide and gradual over long periods of time rather than episodic and locally confined (Ivanov and Head, 2006). Whether there is present-day volcanic activity on Venus is unclear. Some sulfuric clouds have been interpreted as being evidence of recent volcanism (Fegley and Prinn, 1989). Recent data from the IR spectrometer VIRTIS on Venus Express (Helbert et al., 2006; Smrekar et al., 2010) suggest differences in surface temperature (emissivity) that can be linked to morphological features on the surface, indicating changes that are more rapid than the global weathering processes (10^5–10^6 years).

The significant correlation between the geoid and topography at low degrees, together with a relatively high admittance ratio, may indicate that a significant portion of the long-wavelength topography has a dynamic origin by mantle convection (Pauer et al., 2006; Simons et al., 1997). However, it should be noted that, as numerical investigation of convection in strongly temperature-dependent viscosity fluids has progressed, it has become apparent that temperature variations in the nearly stagnant conducting lid are the primary source of topography (Moore and Schubert, 1997; Solomatov and Moresi, 1996) and not dynamic stresses, which dominate at low Rayleigh number and small-viscosity contrast. Thus, topographic rises are not so much held up by upwellings as they are buoyed by the heat delivered to the lithosphere. Plume activity has been suggested by Moore and Schubert (1997), using a model of thermal isostasy. The authors found a very thick thermal lithosphere (200–400 km), which is locally thinned by up to a factor of 5 beneath the volcanic rises. This was attributed to convection in Venus' mantle at a Rayleigh number of about 10^7 and a large viscosity contrast based on the dynamic calculations of Solomatov and Moresi (1996). Similar conclusions are given by Phillips (1994) and Kucinskas and Turcotte (1994).

The influence that the resurfacing event must have had on the thermal history of Venus depends on the resurfacing mechanism. In comparison to the Earth, the question of early plate tectonics on Venus, which has been suggested by several papers, is particularly interesting. But unlike for Mars, there is little present evidence for it. However, it is likely that the main topographic features of plate tectonics, that is, ridges, trenches, and transform faults, may disappear during the 500 Ma after plate tectonics has stopped. Plate tectonics on Earth can take place because the driving forces, that is, slab pull and ridge push, exceed the resistive forces, that is, fault friction, flexure, and viscous drag (Forsyth and Uyeda, 1975). It is commonly suggested that water on the surface of Earth reduces the resistive forces by modifying the rheology of the cold plates such that subduction becomes feasible. The high D/H ratio in the Venusian atmosphere suggests that Venus had substantial water in the past (Fegley, 2004). Today, Venus is desiccated as far as the available spacecraft data can tell. Thus, present plate tectonics are not feasible if water on the surface is required (Nimmo and McKenzie, 1998). Another possible explanation for the present-day lack of plate tectonics on Venus is that the high surface temperature and thick basaltic crust make the lithosphere less dense than the underlying convecting mantle – subduction is inhibited (Herrick, 1994). Alternatively, the high surface temperature may reduce lithospheric stress by reducing the viscosity throughout it (Lenardic et al., 2008; Noack et al., 2012). Furthermore, a high mantle viscosity might cause a large viscous drag force on any moving plates (McKenzie, 1977). On Earth, this mantle drag is greatly reduced by the effect of the low-viscosity zone. For Venus, it has been suggested on the basis of the interpretation of gravity and topography data that a low-viscosity zone beneath the lithosphere does not exist (Kiefer et al., 1986; Nimmo and McKenzie, 1996). Recent models by Pauer et al. (2006) and Steinberger et al. (2010) show that the best fit to the data implies a gradual increase in viscosity.

Another important aspect of the thermal evolution of Venus is the coupling of the atmosphere or hydrosphere with the dynamics of the solid interior. The thick CO_2 atmosphere of Venus results in enhanced surface temperatures due to the greenhouse effect. The surface temperature, which is the upper thermal boundary condition of the convecting interior, on the other hand, influences the heat transport and thus the thermal evolution. These effects have been neglected so far in most

evolution models. A first attempt has been made by Phillips et al. (2001a) who investigated a coupled evolution for Venus using a simple parameterized mantle convection model (including partial melting) and a radiative–convective atmospheric model. Feedback was incorporated by the release of water associated with melt extraction into the Venusian atmosphere, which adds to the greenhouse effect and thus affects the surface temperature. After 2 billion years of evolution, the coupled model was hotter by several tens of kelvins and its extrusive volcanic flux was four times higher than the case without coupling. Their study showed a complex interplay between Venus convective evolution, volcanic activity, and atmospheric state even with such a simple model featuring only basic processes. More complex and elaborate models have been constructed (Lenardic et al., 2008, Noack et al., 2012, Smrekar and Sotin, 2012), demonstrating that the Venusian atmosphere and interior behave as a coupled system and must be considered as such.

Volcanic features on Venus include several large highlands topped by massive volcanoes, a large number of comparatively small shield volcanoes that are apparently randomly distributed, volcanic plains, and about 500 coronae – circular large-scale volcanic–tectonic features. It is widely accepted that coronae form above upwelling plumes (Smrekar and Stofan, 1999). Johnson and Richards (2003) reanalyzed the coronae distribution and had concluded that corona activity is concentrated in the Beta, Atla, and Themis rift systems.

The thickness of the crust has been constrained using gravity and topography data and estimates of the average thickness are about 35 km (Konopliv and Sjogren, 1994; Simons et al., 1994). Thermal evolution models with parameterized convection suggest that Venus may have a 100 km thick lithosphere, which may consist largely of basalt (Spohn, 1991), which has been confirmed by 2-D dynamic models (Armann and Tackley, 2012). This is inconsistent with the gravity results. Those parameterized models, however, assume one-plate tectonics throughout the entire evolution; that is, the crust was not recycled due to either plate tectonics or some other mechanism at the time of resurfacing. Furthermore, as Turcotte (1989) had noted, the thickness of the Venusian crust may be limited by the basalt to eclogite phase transformation that occurs at depths of 30–70 km, depending on the temperature (Ito and Kennedy, 1971; Ringwood, 1975). The dense eclogite phase is gravitationally unstable with respect to the mantle below. When the crust is sufficiently thick, crustal delamination may occur. Crustal thickness variations exist between the various geologic provinces: estimates are between 20 and 50 km at the volcanic rises (Phillips, 1994; Smrekar, 1994), 20–40 km in the plains (Grimm 1994; Konopliv and Sjogren, 1994), and 35–60 km around tessera sites (Grimm, 1994; Kucinskas and Turcotte, 1994; Moore and Schubert, 1997; Simons et al., 1994). The elastic thickness of the lithosphere ranges between 10 and 50 km (Brown and Grimm, 1996; Johnson and Sandwell, 1994; McKenzie, 1994, McKenzie and Nimmo, 1997; Phillips, 1994; Sandwell and Schubert, 1992; Smrekar, 1994), with a clustering around 30 km. Anderson and Smrekar (2006) demonstrated that models that include bottom loading show a larger range of local scales in the crustal thickness and the elastic lithosphere thickness than previous models. The results indicate that geologic processes may be very complex on Venus.

Since the first flyby of Venus by Mariner 2 in 1962 at a distance of 6.6 planetary radii, it was clear that Venus does not have an earthlike magnetosphere. In fact, the following mission of Mariner 5, in 1967, placed an upper limit on the magnetic dipole moment of Venus of $\sim 10^{-3}$ times that of Earth. Later, most definitive measurements of the magnetic moment of Venus were obtained during the Pioneer Venus Orbiter mission in its first years of operation (1979–81). The upper limit on the dipole moment obtained from the Pioneer Venus Orbiter placed the Venus intrinsic magnetic field at less than about 10^{-5} times that of Earth (Russell et al., 1979a,b).

The lack of a present dynamo does not imply that Venus never had an intrinsic magnetic field although we have no information that relates directly to the past history of the field. The surface temperature of about 740 K is relatively close to the Curie temperature (T_{curie}) of the main magnetic carriers on terrestrial planets, such as magnetite, hematite, and pyrrhotite with T_{curie} of 850, 940, and 600 K, respectively (e.g., Dunlop and Özdemir, 1997). Thus, temperatures of the crust – except possibly in an upper thin layer – are expected to be higher than the Curie point (at temperatures below the Curie point, magnetization could persist in rocky materials). Any remanent crustal magnetic field from an early period of dynamo activity might be weak on Venus if existent at all. It is also important to note that the slow rotation of Venus (a Venus day of $\sim$243 Earth days is almost equal to the length of its year of $\sim$224 days, and its sense of rotation is retrograde) does not exclude dynamo action.

Thermal evolution models suggest that there was a magnetic moment of Venus of the same order as the Earth's for about the first 3 billion years of Venus' history (Stevenson et al., 1983; see **Figure 7**). During that time, thermal convection drove the dynamo as suggested for the other terrestrial

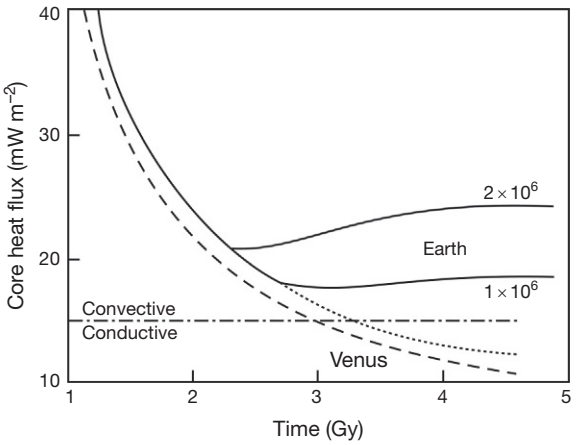


Figure 7 The heat flow out of the cores of Earth (solid and dashed lines) and Venus (dotted line) as a function of time. The curve parameter for the models is the specific energy release upon inner core freezing in J kg^{-1}. The dash-dotted line marks the conducted heat flux along the core adiabat. Without inner core solidification, larger heat fluxes indicate thermal convection in the core. Smaller heat fluxes are then indicative of cessation of thermal convection and dynamo generation. With inner core growth, chemical buoyancy helps to drive convection and the core heat flux may be sub- or superadiabatic. Reproduced from Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.

planets. After the core cooled below the adiabatic heat flow of the core, the magnetic field ceased and did not rejuvenate; the core was too hot and the pressure is too low for inner core growth. The evolution models by Stevenson et al. (1983) have been calculated with the parameterization based on constant viscosity laws, which are more representative for planets with lithosphere delamination (see **Sections 10.08.4.1** and **10.08.4.2**). Assuming that Venus was in a stagnant lid regime throughout the entire evolution, the phase of early dynamo action would be significantly reduced because core cooling might be overestimated with the former models.

An alternative scenario for the present lack of a Venus magnetic field suggests that Venus' field ceased as Venus transitioned from a plate tectonic to a stagnant lid regime around 700 Ma ago (Stevenson, 2002). As suggested for Mars (Nimmo and Stevenson, 2000), the transition heated the mantle and the core. If Venus had an inner core but a mantle temperature increasing in time, then the inner core would have ceased to grow (it may actually be shrinking in radius) and the dynamo could not be driven by chemical buoyancy released upon inner core growth.

10.08.6.3 The Moon

The thermal evolution of the Moon has been a controversial subject. Early models start the evolution of the Moon with an accretionary initial temperature profile. This temperature profile increases from a relatively cold deep interior (less than 1000 K) to reach a maximum in the upper mantle from where temperature decreases toward the surface (Matsui and Abe, 1986). Similar temperature profiles have been used in other models of the Moon's thermal evolution (Cassen and Reynolds, 1973; Cassen et al., 1979; Toksöz et al., 1978). Some of these authors have assumed a magma ocean overlying a cold interior. It has been argued that the absence of global-scale extensional or compressional features on the lunar surface limits variations in the lunar radius to ± 1 km over the past 3.8 Gy (Solomon and Chaiken, 1976). A recent analysis of the Lunar Reconnaissance Orbiter further identified undetected landforms that indicate even recent global shrinking of the Moon (Watters et al., 2010). Solomon and Chaiken (1976) found that for this radius constraint to be satisfied, the contraction due to cooling of the outermost layers of the Moon, which they assumed to be the crystallized product of the magma ocean, must be balanced by radioactive heating of the deeper interior. They argued that this constraint limits magma ocean depths to less than 400 km. Kirk and Stevenson (1989) indicated that a 630 km deep magma ocean can be allowed if contractional cooling of the magma ocean is partially offset by volume changes due to partial melting and chemical differentiation.

There is evidence, however, from the trace element pattern and the remanent magnetization of the Moon's surface rock that the Moon started hot or that the interior heated rapidly after accretion (Binder, 1986; Palme et al., 1984; Runcorn, 1977). The mechanism by which the lunar interior may have been rapidly heated has been the subject of debate. Among the possible mechanisms that have been discussed are the decay of short-lived radionuclides such as ^{26}Al, tidal heating, and electromagnetic heating, none of which are very satisfying because they did not provide heat at the right time or enough heat. The hypothesis of a giant impact origin of the Moon, however, provides a way of accreting the Moon hot (Stevenson, 1987) if the Moon forms more rapidly than the vapor that is ejected from the Earth cools.

More recent evolution models assume an initially hot interior but still open are questions about the thermal and structural state of the Moon after freezing of the magma ocean and whether convection is an important form of heat transport in the lunar interior. It is generally accepted that the anorthositic crust of the Moon formed as a floating crust on a magma ocean. Estimates of the depth of the primordial magmasphere depend on the assumed Al_2O_3 content in the primordial mantle and the assumption that either all Al_2O_3 is incorporated into the plagioclase crust (e.g., Warren, 1985) or some are crystallized also in the Al-bearing Mg–Fe minerals and remained in interstitial melts (Elkins-Tanton et al., 2011). The depth then varies approximately between 400 and 1000 km. Note that the crustal thickness has been recently revised to between 34 and 43 km (Wieczorek et al., 2013); before GRAIL, the average thickness of the Moon's crust was thought to be close to 50 km (Ishihara et al., 2009; Wieczorek et al., 2006). The freezing time of the magma ocean is estimated to last several hundred million years (e.g., Meyer et al., 2010; Minear and Fletcher, 1978; Shearer et al., 2006; Solomon and Longhi, 1977) also to explain the wide range of anorthositic crustal ages (e.g., Borg et al., 2002). In fact, the formation of the anorthositic crust, which acts as an insulating layer for the freezing magma ocean, together with tidal heating from the Earth may be responsible for the respective long time of solidification (Meyer et al., 2010) in contrast to a planet where magma remains at the surface during the solidification process (Elkins-Tanton, 2008). It is assumed that the magma ocean most likely became chemically stratified following its differentiation and crystallization with the late dense iron-rich phases coming to rest upon a less dense Mg-rich phase (apart from the floating anorthositic crust) (e.g., Hess and Parmentier, 1995). Petrologic studies (e.g., Snyder et al., 1992) have shown that this layering will consist of thick layers of olivine and orthopyroxene at the bottom, after which plagioclase crystallizes and floats. One of the last minerals to crystallize is the high-density ilmentite with pyroxene to form an ilmenite-rich layer and, finally, the so-called KREEP layer that is enriched in incompatible and heat-producing elements and sandwiched between the crust and the ilmenite-rich layer, leaving the mantle strongly depleted in radioactive heat sources (**Figure 8**).

Such a layering has strong implications for the subsequent evolution of the Moon since it is prone to convective overturn and mixing (Alley and Parmentier, 1998; Hess and Parmentier, 1995). It has also been speculated that during most of the Moon's history (i.e., after the first rapid overturn), heat might be transported by conduction alone (Kirk and Stevenson, 1989; Pritchard and Stevenson, 1999; Wieczorek and Phillips, 2000) as a consequence of the strong depletion of radioactive elements and the stable chemical layering in the mantle.

Whether the rapid convective overturn scenario took place and whether a subsequent stable chemical layering developed depend on the initial depth of the magma ocean and the freezing mechanism of the magma ocean. For instance,

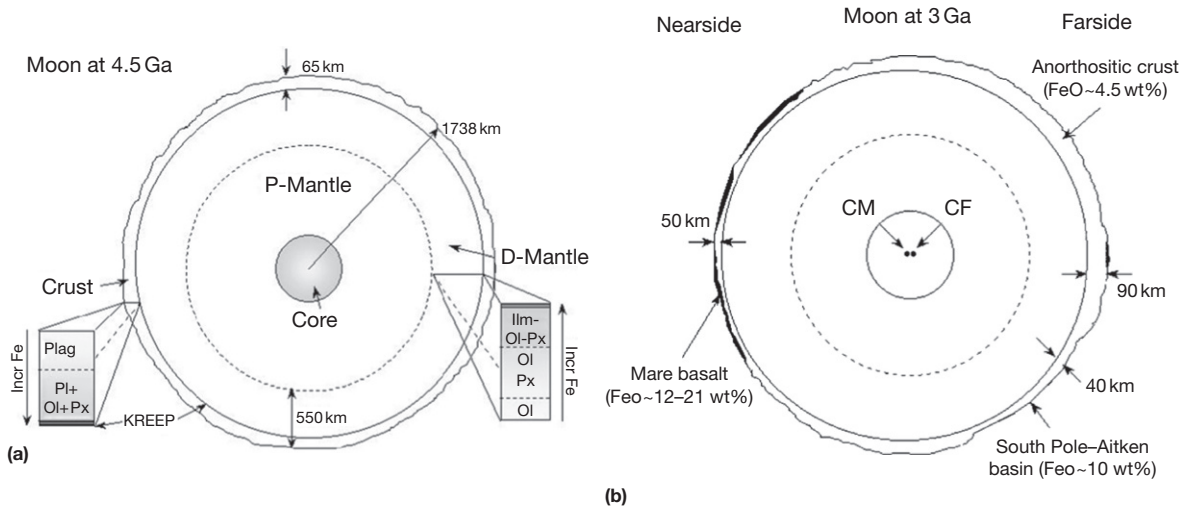


Figure 8 (a) A schematic cross section (roughly to scale) through the Moon immediately after crystallization of the magmasphere. The dashed line represents the 550 km seismic discontinuity. P-mantle is primitive mantle (unmelted or partially melted). 3-D mantle represents differentiated mantle. The sequence of cumulates in the upper mantle shows an upward enrichment in iron with a KREEP layer at the top sandwiched between the crust and mantle. The crustal cumulates are presumed to become more mafic and more iron-rich with increasing depth. (b) A schematic cross section through the Moon at 3 Ga. Note the thick farside crust and the giant South Pole–Aitken basin. CM, center of mass; CF, center of figure. The thickness of mare basalts on the nearside is exaggerated. The FeO concentrations are based on Clementine data (Lucey et al., 2000). Reproduced from McCallum IS (2001) A new view of the Moon in light of data from Clementine and Prospector missions. *Earth, Moon and Planets* 85–86: 253–269.

Elkins-Tanton et al. (2002) showed that the sinking of a solid dense ilmenite cumulate is unlikely because the temperatures at which the cumulates solidify is low, and thus, the viscosity is very high, keeping this layer close to the surface. A possibility to allow this layer to sink, however, is the entrainment of olivine during crystallization. On the other hand, Spohn et al. (2001) showed with 2-D and 3-D convection models that the deep layers of a 700 km deep magma ocean were kept (partially) molten over a significant part of lunar history. It is not well understood whether an unstable chemical layering would develop under these conditions at all. Furthermore, it is likely that the deep layers became chemically rejuvenated by the convective mass flow. The situation is different for a shallow magma ocean. In the case of a 300 km deep magma ocean, a melt zone forms below the magma ocean's original depth in the previously solid residual and convecting mantle; the layering of the freezing magma ocean would most likely not be disturbed but it is even more uncertain whether the unstable dense layers are thick enough to initiate a rapid overturn. In the case of thermal convection of the lunar interior, the evolution is found to be characterized by the growth of a massive 700–800 km thick lithosphere while the lower mantle and core cool only by a few hundred kelvins (Spohn et al., 2001); this is the typical evolution of a one-plate planet.

Based on the different assumptions how the crystallizing magma ocean evolved, the models of the thermochemical evolution of the Moon can be distinguished depending on the assumed initial chemical structure of the mantle. One group (Konrad and Spohn 1997; Laneuville et al., 2013; Spohn et al., 2001; Ziethe et al., 2009) considers a mantle with uniformly distributed composition and heat-producing elements. The mantle is, however, depleted in heat sources as these are enriched in a KREEP layer and distributed either homogeneously (Konrad and Spohn, 1997; Spohn et al., 2001; Ziethe et al., 2009) or laterally inhomogeneous (Laneuville et al., 2013) below the crust. Another group (de Vries et al., 2010; Parmentier et al., 2002; Stegman et al., 2003; Zhang et al., 2013; Zhong et al., 2000) considers a compositionally stratified mantle including an inhomogeneous distribution of the radioactive heat sources as a consequence of magma ocean cumulate overturn. The latter is driven by the dense, ilmenite-bearing cumulates that sink into the mantle and perhaps to the CMB (e.g., Hess and Parmentier, 1995). Both model approaches can not only explain some observations but also have difficulties that will be described in more detail in the succeeding text.

The Moon has been divided into two major geologic terranes, the highlands and the maria. The highlands are more cratered and higher in elevation than the darker maria. The bulk of the highland crustal rocks, consisting of ferroan anorthosites, is formed as a floating crust on a magma ocean during the first 100–200 Ma (Lee et al., 1997). The maria are impact craters filled by basaltic lava. The basaltic lava, the most prominent witnesses of volcanic activity on the Moon, flooded some – but not all – of the major impact basins some 100 Ma after impact basin formation (**Figure 8**). The ages of large mare basalt samples dated radiometrically range from 3.9 to 3.2 Ga. However, the onset of basaltic volcanism appears to be earlier; an age of 4.2 Ga has been reported for a basalt clast in a highland breccia (Taylor et al., 1983). Furthermore, data from the recent GRAIL mission identified large-scale subsurface linear gravity anomalies (Andrews-Hanna et al., 2013). These anomalies are interpreted as intrusive dikes formed before the early Nectarian (~3.8 Ga). In addition to the

possible identification of early volcanism, the structures are thought to result from a global expansion with a lunar radius increase of 0.6–4.8 km. It is however possible that these structures have been formed already during the solidification phase of the magma ocean. Crater counting methods indicate basalts in some unsampled areas of the Procellarum basin are as young as 1.3 Ga (Hiesinger et al., 2000). Such a long-standing volcanic activity can, for instance, be explained by the thermal blanketing effect of the regolith layer (Ziethe et al., 2009).

There is geochemical evidence of a deep origin for the mare basalts having relatively primitive compositions (Hess, 2000). This has often been cited as evidence for deep remelting of the lunar mantle. Furthermore, the TiO_2 content decreases with increasing depth and with increasing age (Basaltic Volcanism Study Project 1981; Head and Hiesinger, 1999), suggesting that the mare basalt source region moves deeper with increasing time. This observation has been confirmed with thermal evolution models using 2-D and 3-D mantle convection (Laneuville et al., 2013; Spohn et al., 2001). The results show a melt zone freezing from above as heat is removed through the growing lithosphere (**Figure 9(b)**) The freezing of the melt zone from above suggests that the magma source region moves to greater depths with time.

The time delay between mare formation and mare fill of about 100 Ma is not well understood. This time delay may be due to the time necessary for the melt to reach the surface after the impact since some time is required for a magma diapir or dike to penetrate through the lithosphere. Furthermore, ascending mantle diapirs likely stall at a basal crustal neutral buoyancy zone (Head and Hiesinger, 1999) before reservoir overpressurization propagates dikes to the surface. Alternatively, it has been suggested that the time delay is the consequence of a chemically stratified mantle following magma ocean differentiation and crystallization (Alley and Parmentier, 1998; Hess and Parmentier, 1995; Zhang et al., 2013). In the presence of a stabilizing compositional density gradient, the onset of thermal convection (due to heating of the interior, the thermal buoyancy can increase with time and overcome the stabilizing compositional density gradient) and thus the volcanic activity will be delayed if the model is started from a nonconvecting state (Alley and Parmentier, 1998). However, if it is assumed that the onset of convection was delayed until tens of Ma after the formation of the maria, the observed volcanic activity prior to the period of mare fill will be left unexplained.

Another open issue of the volcanic activity on the Moon is the observation that the mare basalts are mainly located on the nearside, the side facing the Earth (**Figure 9**). Thus, like Mars, the Moon shows a hemispheric asymmetry that is also reflected in a variation in the crustal thickness: the nearside with most of the mare basalts shows a significantly thinner crust in comparison to the crust on the farside. This was first inferred from the observed 2 km offset between the Moon's center of mass and center of figure roughly along the Earth–Moon axis (Kaula et al., 1972; see **Figure 8(b)**). The nearside–farside dichotomy of crustal thickness has been confirmed from Clementine and GRAIL data (Nozette et al., 1994; Wieczorek et al., 2013). One explanation for the existence of mare basalts in the thinner crust of the nearside is that the hydrostatic pressure of the mare source controls the eruption of mare basalts. Assuming that the maximum depth of the mare source was globally uniform, then mare basalts would only be able to erupt at the surface below a critical elevation. Thus, the higher elevations of the

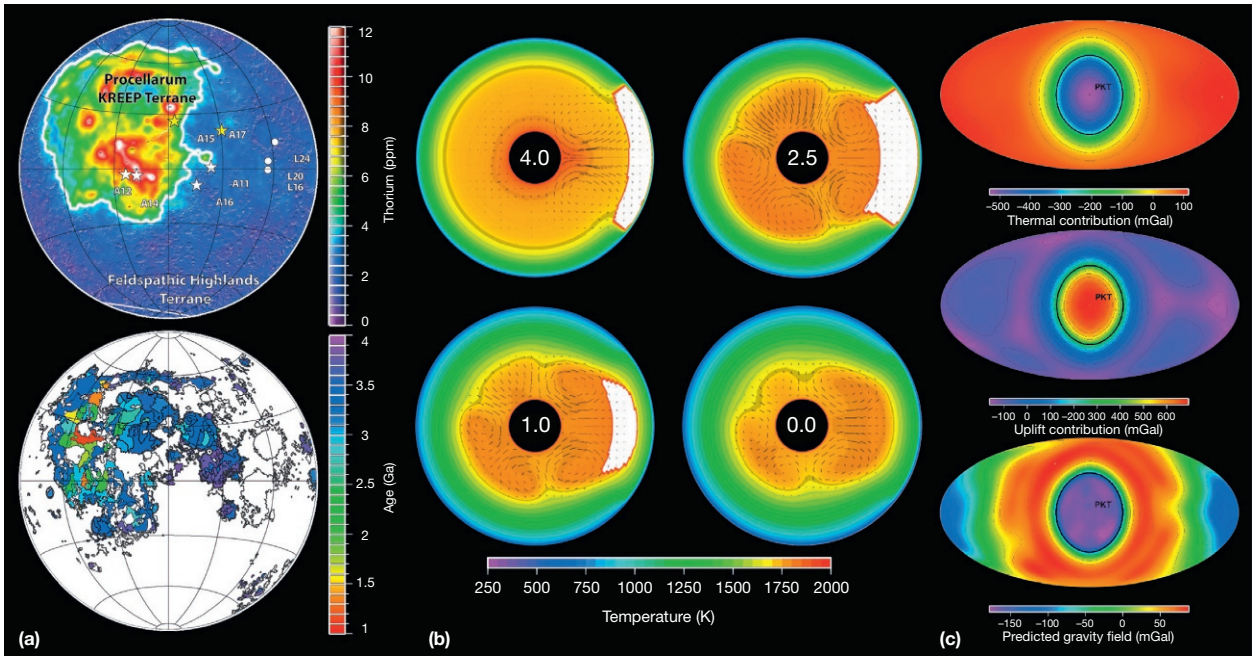


Figure 9 (a) Surface thorium concentration on the nearside (Lawrence et al., 2003): (top) the PKT region outlined in white and (bottom) mare basalt ages from Hiesinger et al. (2003). (b) Evolution of mantle temperature at four different times in Ga for a model with the KREEP layer located below the crust (white: partial melting region). (c) Contribution of thermal anomaly (top) and surface uplift (middle) to the present-day gravity anomaly over the PKT region under the assumption of a strengthless lithosphere (bottom). Reproduced from Laneuville M, Wieczorek M, Breuer D, and Tosi N (2013) Asymmetric thermal evolution of the Moon. *Journal of Geophysical Research* 118: 1435–1452. http://dx.doi.org/10.1002/jgre.20103.

lunar farside could have prevented farside magmas from reaching the surface due to their lack of the necessary pressure in their source. This scenario might, however, be a bit simplistic since the South Pole–Aitken basin has the lowest elevations on the Moon but mare flows in this basin are volumetrically insignificant when compared to the nearside basins.

It has been suggested that the basic lunar asymmetry was established during the magma ocean stage (Loper and Werner, 2000; Ohtake et al., 2012; Warren and Wasson, 1979) since the dynamically unstable configuration of the fractionated magma ocean might result in a large-scale overturn event (Hess and Parmentier, 1995). In that event, part of the KREEP layer is carried away together with the dense ilmenite layer. The large-scale downwelling eventually focused on the nearside would recycle the evolved KREEP-rich material into the deeper mantle providing a heat source for subsequent internal remelting on the nearside (Parmentier et al., 2002; see **Figure 10**). For a large-scale downwelling to develop, the viscosity of the dense layer needs to be significantly lower than the viscosity of the mantle below. Such scenario is difficult to obtain as this layer is expected to be rich in clinopyroxene while the underlying mantle may contain significant amounts of olivine. The latter being more deformable at the same conditions, the ilmenite-enriched layer may be more viscous than an underlying, more olivine-rich mantle at the same temperature (Elkins-Tanton et al., 2002). Alternatively, a large single upwelling may have evolved after the overturn (Zhang et al., 2013; Zhong et al., 2000); the dense ilmenite cumulates together with KREEP are distributed homogeneously atop the CMB but become thermally unstable with time to generate a single upwelling and thus generate the mare basalts (**Figure 10**). To produce a single upwelling, either a small core of less than 300 km (Zhong et al., 2000) – smaller than assumed from seismic observations (Garcia et al., 2011; Weber et al., 2011) – or a low activation energy (Zhang et al., 2013) is required. In addition, the timing, lateral extent, and even existence of the postulated overturn event are not well constrained.

Based on recent gamma-ray data from the Lunar Prospector, it has been realized that the crustal dichotomy is in fact more restricted (Lawrence et al., 1998, 2003). About 60% of the Moon's mare basalts reside in the Procellarum KREEP Terrane (PKT) (Wieczorek and Phillips, 2000). This province is in addition a unique geochemical crustal province enriched in incompatible and heat-producing elements; however, the titanium-rich mare basalts lack the KREEP signature (**Figure 9**). The later finding argues against the scenario, suggesting the rise of KREEP and ilmenite-rich layer from the CMB to produce the mare basalts (de Vries et al., 2010; Parmentier et al., 2002; Stegman et al., 2003; Zhang et al., 2013; Zhong et al., 2000). Wieczorek and Phillips (2000) had argued for a different genetic relationship between the mare basalts and the PKT. Specifically, by distributing a layer of KREEP basalt inhomogeneously below the crust, their thermal, purely conductive model predicts that the lunar mantle should have partially melted only beneath this province. The eruption of mare basalts thus may be primarily controlled by the distribution of heat sources in the Moon. This finding has been recently confirmed with 2-D and 3-D convection models in which an enriched concentration of heat sources in the PKT crust and an otherwise homogeneous mantle have been considered (Laneuville et al., 2013; **Figure 9**). Their model further shows that such an enriched concentration of heat sources in the PKT crust has an influence down to the CMB and leaves a present-day temperature anomaly within the nearside mantle. The latter would predict moderate gravitational and topographic anomalies that are however not observed but may be masked by either crustal thinning or gravitational anomalies from dense material in the underlying mantle.

Paleomagnetic data, combined with radiometric ages of Apollo samples, suggest that a field of possibly 100 μT existed between 3.8 and 3.6 Ga ago (Cisowski and Fuller, 1986; Stephenson et al., 1975). Recent paleointensity measurements on selected lunar samples (e.g., Lawrence et al., 2008) have indicated that lunar rocks are poor magnetic recorders and many of the early estimates of the paleointensities should be considered with caution. The new data, however, suggest a core dynamo may have existed on the Moon from at least 4.2 to 3.2 Ga with a peak in intensity of more than 10 μT at around

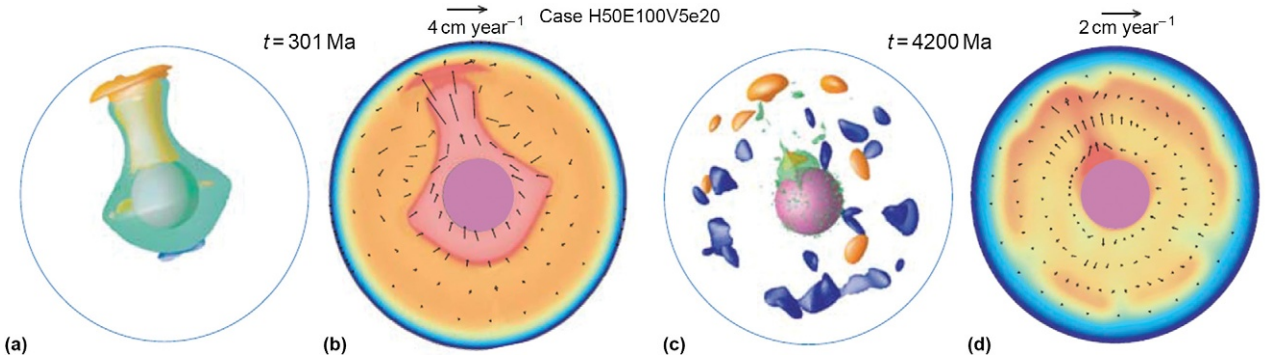


Figure 10 The 3-D thermochemical structures and 2-D cross sections for a model with a ilmenite-bearing cumulate (IBC) layer at the CMB that is enriched in radioactive elements and an activation energy of the mantle material of 100 kJ mol^{-1} (a and b) after 301 Ma and (c and d) at the present day is assumed. The arrows show the flow. Most of the 3-D thermal structures are plotted as isosurfaces of residual temperature with contour levels of −0.07 (blue) and 0.07 (orange) except for (a) that uses the contour level of −0.15 (blue) and 0.15 (orange). The residual temperature is the deviation from the azimuthally average temperature at given depth. The chemical structure is plotted as isosurfaces of compositional field (green). The surface and core–mantle boundary (CMB) are plotted as a blue curve and red spherical surface. Reproduced from Zhang N, Parmentier EM, and Liang Y (2013) A 3D numerical study of the thermal evolution of the Moon after cumulate mantle overturn: The importance of rheology and core solidification. *Journal of Geophysical Research* 118: 1789–1804. http://dx.doi.org/10.1029/jgre.20121.

3.7 Ga (Garrick-Bethell et al., 2009; Shea et al., 2012; Suavet et al., 2013; Tikoo et al., 2012). It should be noted though that due to the lack of data, it is uncertain whether the field ceased 3.2 Ga ago or whether the surface field was just weaker than 1 μT for a longer time after 3.2 Ga (Tikoo et al., 2012). Because the present magnetic field of the Moon is negligible, Runcorn (1975) had argued that the lunar rocks were magnetized at the time of their origin by a field of internal origin. The easiest explanation for such a field is the operation of a dynamo in an iron-rich lunar core. Some researchers doubt that an internal dynamo is required to explain the magnetization of the Moon and favor an alternative idea; the observed magnetic signature is suggested to be generated in association with large impacts during early history (Hide, 1972; Hood and Huang 1991; Hood and Vickery, 1984; Hood et al., 2001). In this model, magnetization should be concentrated diametrically opposite to major impact basins. Satellite observations from the Lunar Prospector mission show concentrations of crustal magnetization at the antipodes of some but not all of the large impact structures. Other models that might explain the younger magnetic signatures include mechanical stirring at the CMB from either precession (Dwyer et al., 2011) or changes in the lunar rotation rate from large impact events (Le Bars et al., 2011).

Assuming an internal origin of the lunar magnetic field, one must be aware that the paleomagnetic data show a gap of about 300 Ma in the remanent magnetization of the lunar rock. There exist two alternative explanations for this gap: (1) The remanent magnetization of older basaltic rock was destroyed by impact gardening and/or the volcanic activity during the first 300 Ma was very low, and (2) the internal field started about 300 Ma after core formation.

An early magnetic field has been suggested from thermal evolution models incorporating 2-D and 3-D convection with strongly temperature-dependent rheology (Konrad and Spohn, 1997; Laneuville et al., 2013; Spohn et al., 2001). These models show that a thermally driven dynamo might have been active from the beginning of the core formation and lasted a maximum of 500 My, that is, 4 Ga ago, if the core was superheated by about 200 K with respect to the mantle. These thermal evolution models, however, assume that at least the lower mantle is not chemically stratified.

A late onset of dynamo action takes into account the thermal and chemical consequences of the magma ocean (Stegman et al., 2003; Zhang et al., 2013). After crystallization of the magma ocean, dense ilmenite and pyroxene cumulates located just underneath the enriched KREEP layer sink to the deep lunar interior. This mixed enriched layer encircles the lunar core and insulates it from the rest of the mantle, trapping heat in the core and preventing the core from cooling convectively, and also from developing a dynamo. After a period of time, the radiogenic material within this 'thermal blanket' decays and heats the lower layer. Eventually, the material becomes more buoyant than its surroundings and rises toward the surface (**Figure 10**). With the removal of the thermal blanket, the core is then able to convect vigorously to cool itself, and this can produce a short-lived thermal dynamo. An essential element of the model is the timing. The length of time that it takes the thermal blanket to heat up and rise back toward the surface is broadly consistent with two important events in early lunar history – the eruption of the mare basalts onto the lunar surface and the magnetization of lunar rocks. Furthermore, the model may explain the phase of global expansion before the early Nectarian (~3.8 Ga) (Andrews-Hanna et al., 2013). Even so, the model has some shortcomings. First, the model would suggest the enrichment of KREEP in the mare basalts, which has not been observed (Lawrence et al., 1998, 2003); instead the enriched material is located adjacent to the basalts in the PKT. Second, for this blanketing layer to become thermally unstable so that core convection can be initiated, the density difference between the lower layer and the overlying mantle can only be a few tens of kg m^{-3} (deVries et al., 2010). The density of the ilmenite-rich layer is about 300 kg m^{-3} higher than the mantle below (Hess and Parmentier, 1995); thus, a mixture of the ilmenite-rich cumulate with the previously fractionated olivine–pyroxene cumulate is necessary for this process to work (Parmentier et al., 2002). The formation of such a layer is uncertain and it is even unlikely to sink due to unfavorable temperature and viscosity conditions (Elkins-Tanton et al., 2002). Furthermore, to obtain a single upwelling, a weak temperature-dependent Newtonian rheology with an effective activation energy of 100 kJ mol^{-1} is required (Zhang et al., 2013) – a value at the very low end even when considering that an effective activation energy for Newtonian rheology can be estimated by multiplying the actual activation energy of the non-Newtonian rheology by a prefactor of 0.3–0.6 (Christensen, 1984).

A recent interpretation of seismic data suggests that the core may have liquid outer and solid inner regions (Weber et al., 2011). Weber et al. (2011) estimated a 240 km radius solidified inner core within a core of 340 km radius. If true, this places strong constraints on the magnetic field generation in the Moon as a chemical dynamo may become a viable alternative – a possibility that has been neglected in most previous models. In case an inner core has grown during the lunar evolution, it needs to be explained why a dynamo has stopped about 3.2 Ga ago (or at least at the present day) although chemical convection is very efficient for the magnetic field generation (Braginsky, 1964; Stevenson et al., 1983). In addition, as a consequence of the low pressure range in the Moon's core, for a sulfur content higher than about 8 wt%, iron snow (see **Section 10.08.5.2.1**) is a likely mechanism for core solidification as it has been already suggested for Ganymede (e.g., Hauck et al., 2006), Mars (Stewart et al., 2007), and Mercury (Chen et al., 2008).

Recent studies suggest that water is more abundant in the lunar interior than previously thought (e.g., Grimm, 2013; Klima et al., 2013; McCubbin et al., 2011; Saal et al., 2008). Its distribution in the interior is possibly inhomogeneous as a consequence of the solidification of the magma ocean (Elkins-Tanton and Grove, 2011; McCubbin et al., 2011). In any case, the observation of interior water may have profound implications on the thermochemical evolution of the Moon as water even in small concentrations is known to have a strong effect on mantle rheology, with a viscosity reduction of more than a factor of 100 between water-saturated and completely dry material (e.g., Hirth and Kohlstedt, 1996; Mei and Kohlstedt, 2000). So far, most models have considered a dry rheology and the associated temperature profiles are generally consistent with the deep mantle temperatures suggested from seismic and gravitational inversions (e.g., Gagnepain-Beyneix et al., 2006; Khan et al., 2007) and electric conductivity studies

(Grimm, 2013; Hood et al., 1982). In general, a wet rheology predicts more efficient cooling, lower present-day mantle temperatures, and an earlier cessation of volcanism. This on the other hand would also suggest a higher sulfur concentration in the core to explain either an entirely fluid core or a solid inner core of 240 km (Weber et al., 2011).

10.08.6.4 Mars

A fundamental problem in the evolution of Mars is the timing and the origin of the crustal dichotomy. The southern highlands and northern lowlands of Mars differ markedly in average elevation (Smith et al., 1999a,b) and crustal thickness (Neumann et al., 2004; Zuber et al., 2000). Although it is generally accepted that this crustal dichotomy is one of oldest features on Mars, the exact timing of the dichotomy formation, which has implications for the formation mechanism, is strongly debated. Early studies on faulting and other tectonic features at the crustal dichotomy boundary suggested a Late Noachian/Early Hesperian (3.7 Ga) formation for the dichotomy (McGill and Dimitriou, 1990). Most recent studies suggest an Early Noachian or earlier (>3.9 Ga) formation time for the dichotomy (Frey et al., 2002; Nimmo and Tanaka, 2005). One argument for this early crustal formation is the finding of quasi-circular depressions in the northern lowlands (Frey et al., 2002). They are interpreted as buried impact craters and are suggestive that the basement of northern lowlands formed during the Early Noachian and may be even older than the visible highlands. Solomon et al. (2005) even suggested an early crustal formation, including the dichotomy structure, as early as the first 50 Ma after the solar system formation. Lines of evidence include the isotopic anomalies in Martian meteorites (Borg et al., 1997; Brandon et al., 2000; Chen and Wasserburg, 1986; Halliday et al., 2001; Harper et al., 1995) that indicate an early differentiation event of the mantle about 4.5 Ga ago. The SNC meteorite isotope data further suggest insignificant reservoir mixing thereafter (Jagoutz, 1991).

The origin of the crustal dichotomy has variously been related to external (Frey and Schultz, 1988; Wilhelms and Squyres, 1984) and internal processes (Wise et al., 1979), but none of the proposed formation mechanisms has been fully convincing in part due to the uncertainty in the timing of the dichotomy formation. For the external processes, one impact (Wilhelms and Squyres, 1984) or several large impacts (Frey and Schultz, 1988) have been suggested as an explanation of the crustal dichotomy. This idea has recently been revived in a series of articles explaining the observed elliptical shape of the lowlands (Andrews-Hanna et al., 2008; Marinova et al., 2008; Nimmo et al., 2008). However, an endogenic origin of the dichotomy cannot be ruled out. For an endogenic origin of the dichotomy, four different mechanisms have been proposed that are associated with (1) the evolution of an early magma ocean (Elkins-Tanton et al., 2003; Hess and Parmentier, 2001), (2) an episode of degree-one mantle convection (Ke and Solomatov, 2006; Keller and Tackley, 2009; Schubert and Lingenfelter, 1973; Wise et al., 1979; Zhong and Zuber, 2001), (3) a thermal anomaly following a giant impact in the southern hemisphere (Golabek et al., 2011; Reese and Solomatov, 2006; Reese et al., 2010), and (4) an early phase of plate tectonics (Sleep, 1994).

The crustal dichotomy may have arisen from a heterogeneous evolution of a magma ocean or the crystallization of a magma ocean may have led to gravitationally unstable layering (a similar suggestion has been proposed for the Moon, which also exhibits a hemispheric crustal thickness variation; see **Section 10.08.6.3**), as the late-stage silicates that crystallized at shallow mantle depths were denser than earlier cumulates that crystallized near the base of the magma layer. A potential long-wavelength overturn of an unstable mantle may have thickened the crust over the downwelling region and thinned the crust elsewhere (Hess and Parmentier, 2001) although this fails to explain the sharpness of the dichotomy or the relative uniformity of the crustal thickness in the southern highlands.

An episode of long-wavelength mantle convection after the solidification of an early magma ocean has been suggested for a layered mantle viscosity structure with a viscosity jump of more than a factor of 25 in the midmantle (Roberts and Zhong, 2006; Zhong and Zuber, 2001) or alternatively a midmantle viscosity jump linked to the Martian mineralogical transition zone below 1000 km depth (Keller and Tackley, 2009). Another situation that yields degree-1 mantle convection is the possible existence of the endothermic spinel-to-perovskite just above the Martian core (Harder and Christensen, 1996). The time required to develop degree-1 mantle convection ranges from 100 Ma to several hundred Ma and is too slow to satisfy the geochemical constraints on the timing of the dichotomy formation. Additionally, these models result in continuous magmatic resurfacing of the southern hemisphere, thereby destroying the initially dichotomous crustal structure in the long term. Ke and Solomatov (2006) on the hand suggested a large, transient upwelling that arose from a hot bottom thermal boundary layer generated by the initially hot Martian core. A superheated core relative to the mantle of 1000 K is required to form one large plume. Its development can be very fast, on the order of 10–1000 My depending on the value of mantle viscosity. None of these models explains the sharpness of the dichotomy nor the relative uniformity of the crustal thickness in the southern hemisphere.

A hybrid model of exogenic and endogenic origin has been proposed by Reese et al. (2010) and Golabek et al. (2011). A giant impact event during the late stage of planetary formation is the source of the southern highland crust. In a second stage, the extraction of excess heat by vigorous mantle convection on the impacted hemisphere leads to massive magmatism, forming a distinct Tharsis-like volcanic region (Golabek et al., 2011). The location of the volcanic region is, however, inconsistent with the observed position of Tharsis at the dichotomy boundary. A solution to that problem might be the relative rotation of the entire lithosphere relative to the underlying mantle. Taking the presence of a hemispheric lithospheric keel into account that represents a depleted mantle reservoir, Šrámek and Zhong (2012) demonstrated that a stiff keel can induce a rotation, until an upwelling is stabilized near the edge of the keel. This edge corresponds to the dichotomy boundary, thus accounting for the location of the Tharsis rise. Sleep (1994) proposed that the smooth northern lowlands and the Tharsis volcanoes were produced as an ocean floor and an island arc volcanic chain similar to these features on Earth. More recently, the magnetic lineation patterns on parts of the

southern highlands detected by MGS have been interpreted to be the result of plate divergence (Connerney et al., 1999, 2005). The dating of the magnetic anomalies led the authors to suggest that the potential early plate tectonic regime lasted about 500 Ma. However, there is no consistent geologic evidence for such a regime on Mars (Nimmo and Tanaka, 2005).

The proposed scenarios for the formation of the crustal dichotomy have different implications for the thermal evolution of Mars, in particular for the early evolution. At the present stage and most likely during the last 4 Ga, Mars is and was in the stagnant lid regime with a stable plate on top of the convecting mantle. The first few hundred million years, however, is an uncertain epoch and it is unknown whether the planet was in a phase of plate tectonics that later changed to one-plate tectonics or whether the planet never changed its heat transport mechanism and always was in the stagnant lid regime.

The existence of a brief episode of plate tectonics has important consequences for the thermal evolution of the planet. Assuming stagnant lid convection throughout the entire evolution of Mars, most models show a continuously cooling interior. In contrast to the stagnant lid convection throughout the entire evolution, an early plate tectonic regime suggests a rapidly cooling interior during the first few hundred million years. As soon as plate tectonics stops, the heat transport becomes inefficient and the interior heats up until the temperature again decreases slowly to its present value that does not differ significantly to the temperature of the models with stagnant lid convection throughout the entire evolution assuming the same reference viscosity (**Figure 11**). The temperature increase also suggests a decrease in the thickness of the elastic lithosphere after cessation of plate tectonics and during a long

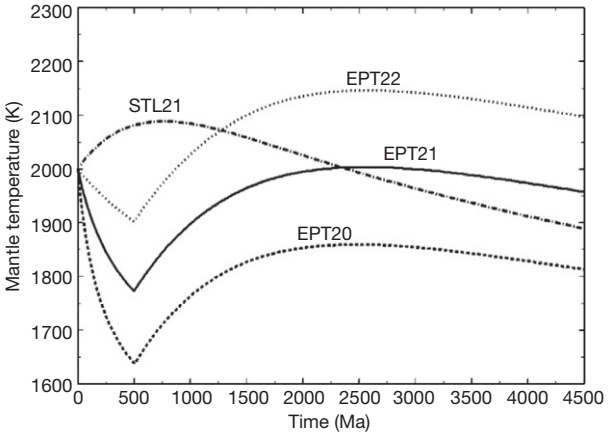


Figure 11 Mantle temperature as a function of time for three models with early plate tectonics. After 500 Ma, plate tectonic ceases and stagnant lid convection sets in. Model EPT20 has a reference viscosity of 10^{20} Pa s (dashed line), model EPT21 a reference viscosity of 10^{21} Pa s (solid line), and model EPT22 a reference viscosity of 10^{22} Pa s (dotted line). Also shown is the temperature for the stagnant lid convection model STL21 with a reference viscosity of 10^{21} Pa s (dash-dotted line). This model assumes stagnant lid convection throughout the entire evolution of Mars. Reproduced from Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research: Planets* 108(E7): 5072. http://dx.doi.org/10.1029/20002JE001999.

period of Martian evolution. Such a trend, however, has not been observed (McGovern et al., 2004; Zuber et al., 2000).

The elastic lithosphere thicknesses, which were derived from an analysis of the gravity and topography data, suggest a rapid growth during the Noachian and a more modest subsequent growth (McGovern et al., 2004; Zuber et al., 2000). This trend is in general consistent with the thermal evolution models (Breuer and Spohn, 2003, 2006; Breuer et al., 1993; Grott and Breuer, 2008; Ruiz et al., 2011; Schubert et al., 1992; Spohn et al., 1998), which show a rapid growth of the elastic lithosphere (defined by an isotherm of 1050 K) from a few kilometers to more than 100 km during the first few hundred million years and later a slower growth to values of 130–200 km. A present-day value of the elastic thickness has been derived using radar sounding data obtained by SHARAD, the shallow radar onboard the Mars Reconnaissance Orbiter; Phillips et al. (2008) found that the Martian lithosphere is extremely stiff and T_e is larger than 300 km at the north pole today. This is almost twice as large as previously estimated from thermal evolution models and contradicts the best fit elastic thickness derived from flexure studies of the south polar load with $T_e = 140$ km, although any value greater than 102 km can fit the observations (Wieczorek, 2008). The large present-day elastic lithosphere can be explained if the planet's bulk heat production elements are about 25–50% subchondritic or – which is more likely – the large elastic thicknesses in excess of 300 km are not representative for the bulk of the planet and T_e possibly shows a large degree of spatial heterogeneity (Grott and Breuer, 2009; Ruiz et al., 2011; **Figure 12**).

In addition, the estimated small values of the effective elastic lithosphere thickness of less than 16 km (Grott et al., 2005, 2007; McGovern et al., 2004; Zuber et al., 2000) are difficult to explain unless the observed features are Early Noachian in age (more than 4 Ga; Hartmann and Neukum, 2001). In particular, the analysis of rift valley topography in the Thaumasia region (Grott et al., 2005, 2007) where the formation of the examined rift zones has actually been dated at the transition of the Noachian to the Hesperian ($\sim$1 Gy after accretion) suggests an effective elastic lithosphere thickness of about 12 km, which is inconsistent with the present models showing a rapid growth of the elastic lithosphere to more than 100 km during the first few hundred million years.

An alternative explanation for the observed rapid growth of the elastic lithosphere thickness is given by the effect of rheological layering of an enriched and low-conductivity primordial crust on the thermal evolution (Grott and Breuer, 2008, 2009; Schumacher and Breuer, 2006). Thus, the rapid growth of the elastic lithosphere can be the consequence of a change of the mechanical state of the crust–mantle interface, with a low elastic thickness corresponding to mechanical decoupling and high elastic thickness corresponding to mechanical coupling of the crust and mantle. Decoupling in the young and warm lithosphere is caused by the presence of an incompetent layer in the lower crust that later disappears due to lithospheric cooling. In order for an incompetent layer to exist, a weak crustal rheology is required (Grott and Breuer, 2008, 2010), and given the absence of large-scale felsic deposits (e.g., Nimmo and Tanaka, 2005), it seems likely that water is responsible for rheological weakening. Furthermore, small elastic

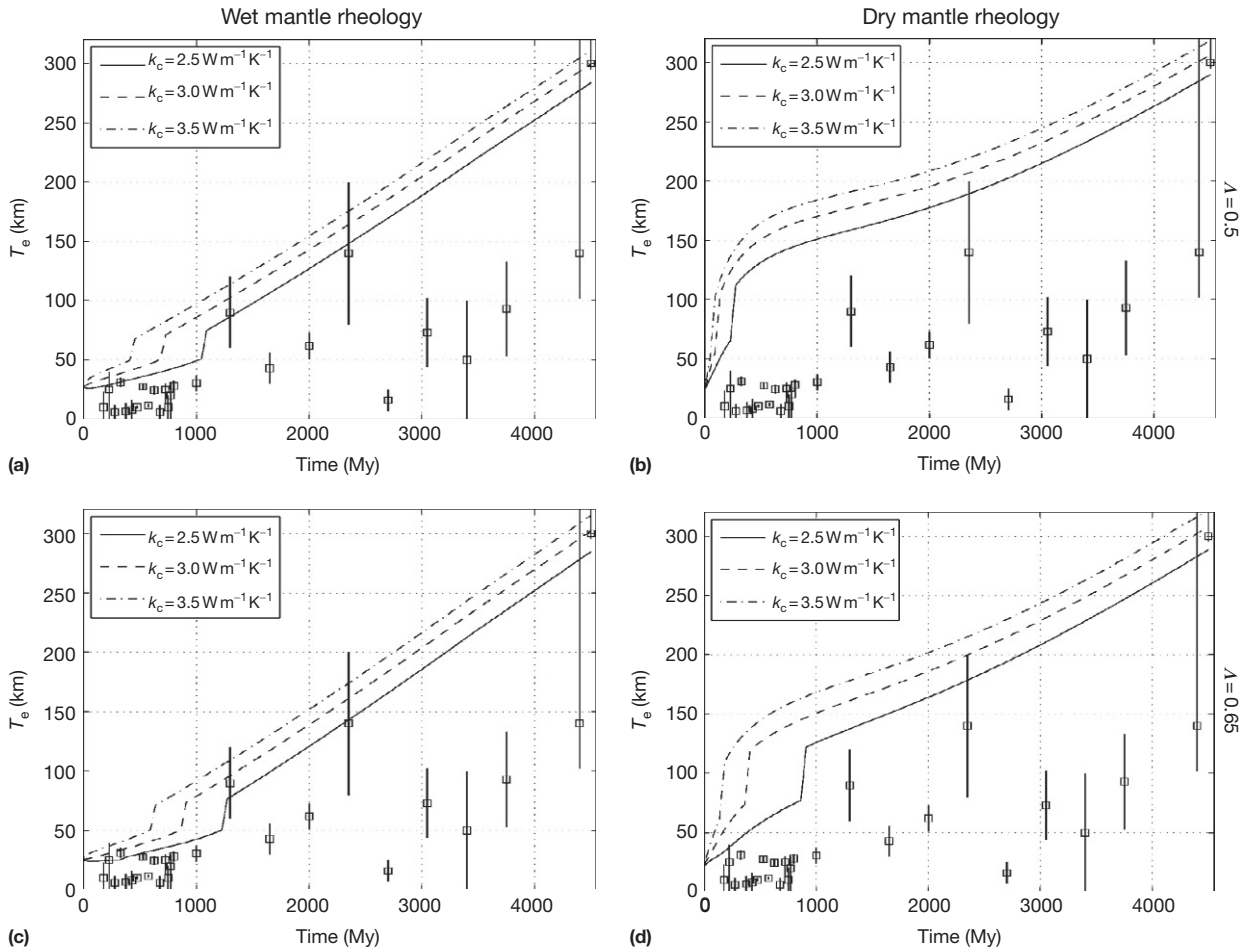


Figure 12 The elastic lithosphere thickness T_e as a function of time for thermal evolution models with (a) and (c) a wet mantle rheology and (b) and (d) a dry mantle rheology, varying the crustal thermal conductivity k_c. The bulk concentration of heat-producing elements has been reduced with respect to the reference model (Wänke and Dreibus, 1994) such that today's elastic thickness is ~300 km (Phillips et al., 2008). Remaining radioactive elements have concentrations of (a) $\varphi = 0.5$, (b) $\varphi = 0.75$, (c) $\varphi = 0.65$, and (d) $\varphi = 1$ with respect to the reference model. Heat-producing elements are enriched in the crust and $\Lambda = 0.5$ and 0.65 in (a)–(b) and (c)–(d), respectively. Elastic thickness estimates from the literature (Belleguic et al., 2005; Grott et al., 2005, 2007; Hoogenboom and Smrekar, 2006; Kronberg et al., 2007; McGovern et al., 2004; Phillips et al., 2008; Schultz and Watters, 2001; Wieczorek, 2008) are also given. Reproduced from Grott M and Breuer D (2009) Implications of large elastic thicknesses for the composition and current thermal state of Mars. *Icarus* 201(2): 540–548.

thicknesses require relatively thin thermal boundary layers in the upper mantle, indicating that mantle convection was active and vigorous during this time period. A vigorous convecting mantle is, however, more difficult to reconcile with the sustainment of early reservoirs in the Marian mantle to explain the SNC isotope characteristic (Jagoutz, 1991; see also discussion later in the text). In fact, thin thermal boundary layers also require low mantle viscosities, indicating that the convecting mantle is best described by a wet or only partially dehydrated rheology (**Figure 12**). A wet crust is consistent with an early, deep-seated global-scale hydrologic system on Mars (Andrews-Hanna et al., 2007; Parmentier and Zuber, 2007), which would be conducive to a wet crustal rheology. The water content of the Martian mantle, however, is strongly debated. The SNC meteorites are one of the few constraints available to estimate the amount of water in the Martian mantle. The bulk water content of the SNC meteorites is relatively small and ranges from 50 to 150 ppm (Leshin et al., 1996), indicating low mantle water contents in the source regions of the SNC meteorites between only 5 and 15 ppm (Filiberto and Treiman, 2009). However, it has been argued that magmatic degassing could be responsible for the low bulk water content in the SNC meteorites (Balta and McSween, 2013; McSween et al., 2001), and the water content of the Martian mantle could be considerably larger with values between 55 and 220 ppm H_2O (McCubbin et al., 2010, 2012b), values similar to that of the Earth's mantle.

Crustal formation provides an important link between mantle and atmospheric reservoirs (see review in Grott et al., 2013) as volcanic outgassing is one of the main sources of volatiles for the Martian atmosphere and it depletes the mantle with time in these elements. Water behaves similar to other moderately incompatible elements and will preferably move to the liquid phase upon partial melting. Thus, it will be extracted from the mantle during the crustal formation process. Part of this water will then be outgassed into the atmosphere through

extrusive volcanism. Based on thermochemical models, the water depletion in the mantle ranges between 50% and 80% (Fraeman and Korenaga, 2010; Grott et al., 2011b; Morschhauser et al., 2011), with 50% being the preferred model that can explain various observables on Mars (Grott et al., 2011b). The latter models imply that the mantle is homogeneously depleted. Models that also consider compositional variation due to melting, however, suggest a depleted upper mantle (depleted in crustal compounds and volatiles) and a lower primordial mantle that still contains a substantial amount of water (Plesa and Breuer, 2014; **Figure 14**) as suggested by a recent analysis of the SNC meteorites (Balta and McSween, 2013).

Mars has been volcanically active throughout its history. In fact, the recent detection of young lava flows (< 100 Ma) suggests that the planet has been volcanically active up to the recent past (Hartmann et al., 1999). A continuous decline of the volcanic activity has been suggested since the Noachian. Martian volcanism and stratigraphy have been reviewed by Greeley and Spudis (1981), Tanaka et al. (1992), and Mouginis-Mark et al. (1992). The time of crustal formation has been constrained using isotope data gathered from SNC meteorites (see Halliday et al., 2001, for a recent review) and surface morphological data on volcanic extrusions. A study of the Nd mass balance in Martian meteorites (Norman, 2002) proposes a two-stage formation of the Martian crust with 20–30 km crust formed early, possibly simultaneously with core formation, and 45–75 km thereafter from depleted mantle sources. The exact timing of the second stage, however, cannot be constrained with the geochemical data.

The present thickness of the Martian crust is also not very well known and estimates of the average thickness vary between 40 and 150 km. The upper value has been derived from models of the interior structure of Mars based on the moment of inertia factor and the chemistry of the SNC meteorites (Sohl and Spohn, 1997; Sohl et al., 2005) and the lower value from a Bouguer inversion of the gravity and topography data assuming Airy isostasy (or a crust density that is everywhere the same) and further assuming that the floor of the Hellas impact basin defines the minimum crustal thickness (Zuber et al., 2000). Combining the different constraints on the crustal thickness, a globally averaged crustal thickness of 50 ± 12 km (Wieczorek and Zuber, 2004) seems likely (see **Chapters 10.02** and **10.05** for more details) in agreement with a crustal thickness recently estimated from geochemical measurements (Humayun et al., 2013).

The thermochemical evolution models that include crustal growth with either parameterized (Breuer and Spohn, 2003, 2006; Breuer et al., 1993; Fraeman and Korenaga, 2010; Grott et al., 2011b; Hauck and Phillips, 2002; Morschhauser et al., 2011; Schubert et al., 1992; Spohn, 1991) or 2-D and 3-D convection models (Golabek et al., 2011; Keller and Tackley, 2009; Li and Kiefer, 2007; Ogawa and Yanagisawa, 2011, 2012; Ruedas et al., 2013a,b; Schumacher and Breuer, 2006; Šrámek and Zhong, 2012) usually predict that most of the crust formed in the first billion years. These models in general confirm the observation of a continuous decline of global volcanism and crustal growth since the Noachian. Furthermore, the initial temperature distribution in the Martian interior and the viscosity structure of the Martian mantle were identified as the key parameters influencing mantle melting and the evolution of the Martian crust – thus the thickness of the crust at the end of the thermal evolution. In contrast, models with early plate tectonics in Mars show an early efficient cooling during the phase of plate tectonics that will either frustrate later crustal production and may not explain, for example, the northern volcanic plains in the Early Hesperian or show a crustal growth rate with a late peak (about 2 Ga) that is not observed (Breuer and Spohn, 2003).

The strict assumption that most of the crust is produced during the first few hundred million years (Nimmo and Tanaka, 2005; Solomon et al., 2005) places a strong constraint on thermal evolution models that assume stagnant lid convection throughout the entire evolution. Hauck and Phillips (2002) argued that only a wet mantle with an initial mantle temperature of about 1700 K can explain this observation. In contrast to these findings, Breuer and Spohn (2006) showed that models with a dry Martian mantle are also consistent with the crustal evolution if the initial mantle temperature after core formation is high enough and if a primordial crust has been formed during core formation. Irrespective of mantle viscosity, the present-day mantle cooling rates are 40–50 KGy^{-1}. These values are in general agreement with geochemical observations from which cooling rates of 30–40 KGy^{-1} have been derived (Baratoux et al., 2011). The ambiguity of a wet versus dry Martian mantle has recently been investigated on the basis of elastic thickness estimates, which indicate that a dry mantle rheology is incompatible with the low elastic thicknesses suggested for the Noachian and Early Hesperian periods (Grott and Breuer, 2008, 2010).

The observed recent volcanic activity in the large volcanic provinces, Tharsis and Elysium (Berman and Hartmann, 2002; Hartmann et al., 1999; Neukum et al., 2004), places another important constraint on the thermochemical evolution. In these regions, the volcanic activity was most likely active for a few billion years although the bulk of the volcanic provinces was present since the Noachian (Banerdt and Golombek, 2000; Phillips et al., 2001b). Thermochemical evolution models (Breuer and Spohn, 2003, 2006; Fraeman and Korenaga, 2010; Hauck and Phillips, 2002) suggest that crustal formation ceased 1–3 billion years ago based on the fact that the average mantle temperature fell below the mantle solidus. Among the models assuming a wet mantle rheology, the crustal production ceases faster due to more efficient cooling. To obtain long-standing melt production, in particular, an insulating crust and the formation of a depleted upper mantle layer help to prolong the volcanic activity even with a weak mantle rheology (Morschhauser et al., 2011; Plesa et al., 2014). In particular, the influence of an insulating crust to prolong the volcanic activity has been already shown for the Moon (Ziethe et al., 2009) and Mercury (Grott et al., 2011a; Tosi et al., 2013). Note, however, that parameterized models cannot account for lateral variations in crustal production, and fully 2-D or 3-D convection models need to be applied in order to resolve young, localized volcanism.

The existence of long-standing volcanism in Tharsis and Elysium up to recent times could be explained with stable and long-standing plumes that have a higher temperature than the average mantle. As both Tharsis and Elysium are located near or at the dichotomy boundary, a dynamic link

between these two features has been suggested (Golabek et al., 2011; Šrámek and Zhong, 2010, 2012; Wenzel et al., 2004; Zhong, 2009). Possible mechanisms to generate stable and long-standing plumes may be due to a deep phase transition (Breuer et al., 1998; Harder, 2000; Harder and Christensen, 1996; Weinstein, 1995). In particular, the endothermic spinel-to-perovskite phase transition that might be located close to the CMB (see also **Chapter 10.02**) helps to stabilize the plumes and reduces their number to one or two upwellings, although the long timescale necessary to form a single upwelling in most models is inconsistent with an early formation of either the dichotomy or the volcanic provinces (Michel and Forni, 2011; Roberts and Zhong, 2006). Another efficient way to induce low-degree convection has been suggested by a pronounced rheological stratification with an increase of the viscosity with depth (Golabek et al., 2011; Keller and Tackley, 2009; Roberts and Zhong, 2006; Šrámek and Zhong, 2010; Zhong, 2009). Presently, it is uncertain whether such pronounced viscosity stratification exists on Mars. A detailed model by Ruedas et al. (2013a,b) in which the viscosity depends on temperature, pressure, melt, and mineralogy does not obtain low-degree convection during the thermal evolution.

Another mechanism to generate long-standing plumes – although not just a single one – has been proposed assuming chemical layering with either a variable mantle thermal conductivity (Schott et al., 2001), a low conducting crust (Plesa and Breuer, 2014), or the presence of the crustal dichotomy (Wenzel et al., 2004). For these mechanisms, the existence of a chemically layered mantle is required to occur early in the evolution due to fractionation of a magma ocean (Wenzel et al., 2004) or as the consequence of crustal formation due to the formation of a depleted upper mantle layer, which does not remix with the remaining undepleted lower mantle (Plesa and Breuer, 2014; Schott et al., 2001). Furthermore, an alternative mechanism has been proposed by Schumacher and Breuer (2006) that may also work in addition to the previous scenarios. Once the main structure of the volcanic provinces has been built up, the lateral thickness variations of the low-conductivity crust allow the continuous generation of melt underneath these thickened regions and may explain the long-standing and recent volcanism in Tharsis and Elysium. It has been suggested that the crust underneath Tharsis is about 20–40 km thicker than in its surroundings (Neumann et al., 2004).

Most thermal evolution models implicitly assume the Martian mantle to be homogeneous, but the existence of distinct geochemical reservoirs is suggested by isotopic anomalies in the Martian meteorites (Borg et al., 1997; Brandon et al., 2000; Chen and Wasserburg, 1986; Halliday et al., 2001; Harper et al., 1995). These data indicate an early mantle differentiation event about 4.5 Ga ago and insignificant reservoir mixing thereafter. One possible mechanism to form distinct and separate geochemical reservoirs is magma ocean solidification by fractional crystallization, which likely results in an unstable density stratification in the Martian mantle. Late-stage silicates that crystallize at shallow depths are denser than earlier cumulates that crystallize near the base of the magma ocean – a situation prone for vigorous mantle overturn (Elkins-Tanton et al., 2003, 2005; Hess and Parmentier, 2001).

The mantle overturn is assumed to occur during the first 100 Ma (Debaille et al., 2007, 2009; Elkins-Tanton et al., 2003, 2005a,b). As a consequence of the overturn, the dense surface material sinks to the CMB while the light material from the CMB rises to the surface – a stable chemical-layered mantle develops (**Figure 13**), which is in its composition in particular in the midmantle laterally heterogeneous (Elkins-Tanton et al., 2005; Zaranek and Parmentier, 2004). It is further assumed that, during the vigorous overturn, rising upwellings can melt by adiabatic decompression and produce an early crust. The residual mantle of this process is, for instance, associated with the source region of the nakhlites (Debaille et al., 2009). The chemical layering after the overturn suppresses subsequent thermally driven convection and early-formed reservoirs are preserved throughout the entire planet's history.

Although such a process would form distinct geochemical reservoirs, models of the interior evolution after mantle overturn indicate that subsequent thermal convection – if existent at all – is very weak. This contradicts the small elastic lithosphere thicknesses derived for the Noachian period, which require the lithosphere to be warm and the upper thermal boundary layer to be thin (Grott and Breuer, 2008). Furthermore, different reservoirs remain dynamically strictly separated, because the density differences between different layers are large (Tosi et al., 2013b; Plesa et al., 2014). A potential reservoir in the deep mantle would have no surface expression and thereby would remain 'invisible,' contrary to observations. Therefore, alternative explanations for the early formation of reservoirs that did not mix significantly during the subsequent evolution are needed. One such model includes the rapid formation of a harzburgite layer (Ogawa and Yanagisawa, 2011, 2012; Plesa and Breuer, 2014; Schott et al., 2001) by early crustal formation, which would be less dense than the primordial mantle and would remain stable, while the lower mantle layer consists of compositionally denser materials not depleted in the basaltic component – the larger the density contrast between depleted mantle and primordial mantle, the more stable is the layering (**Figure 14**). The stability of the layering is even enhanced as a consequence of dehydration and subsequent stiffening in case of a wet mantle (Plesa and Breuer, 2014). These two layers convect separately and hot plumes grow from the lower mantle layer making it thinner with time by erosion – its composition, however, remains almost primordial (**Figure 14**). The stiff and buoyant upper mantle layer thermally isolates the interior, and due to the associated slow cooling of the mantle, partial melting can exist for a substantial time period in the evolution. The longevity of volcanism is further enhanced by the low conducting crust with the upwellings in the depleted upper mantle layer being very stable (Plesa and Breuer, 2014). In the case of crustal recycling, another 'hidden' reservoir consisting of eclogitic material that is denser than the primitive mantle would accumulate at the CMB (Ogawa and Yanagisawa, 2011, 2012). Whether such a model can also explain the isotope characteristic of the SNC meteorites remains to be studied.

At present, Mars has no global dipole magnetic field, but regions of strongly magnetized ancient crust indicate that an internal field once existed, one of the most astonishing outcomes of the MGS mission (Acuña et al., 1998, 1999, 2001; Connerney et al., 1999). After the MGS Orbit Insertion in 1997

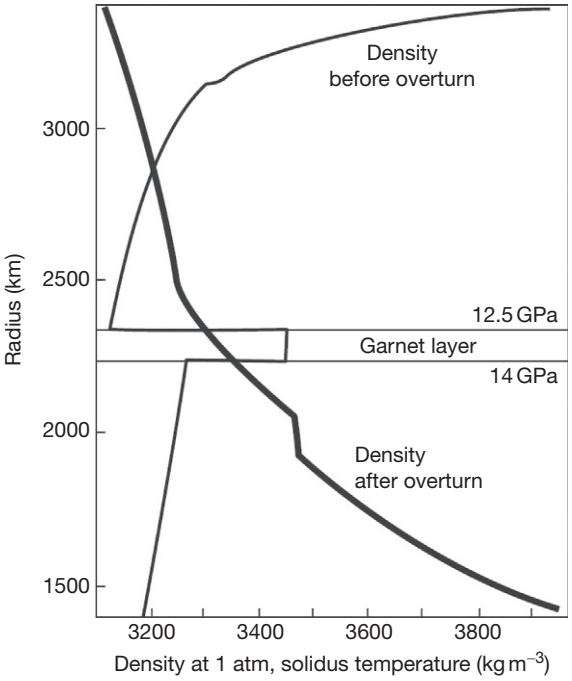


Figure 13 Preoverturn and postoverturn density profiles for a simple fractional crystallization model of a whole mantle Martian magma ocean showing the gravitational stability of the mantle after overturn and the existence of overlapping density gradients before overturn that result in lateral compositional and temperature heterogeneities in the overturned mantle. Reproduced from Elkins-Tanton LT, Hess PC, and Parmentier EM (2005) Possible formation of ancient crust on Mars through magma ocean processes. *Journal of Geophysical Research* 110: E12S01, http://dx.doi.org/10.1029/2005JE002480.

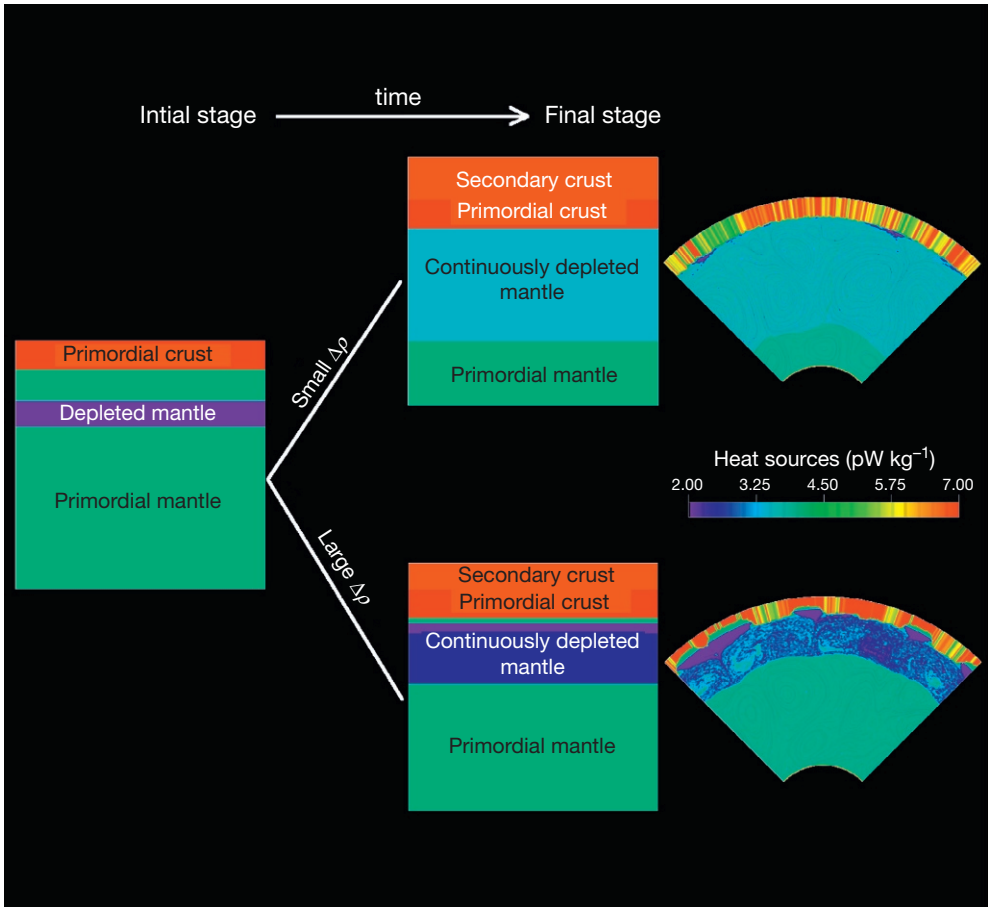


Figure 14 Formation of geochemical reservoirs in the Martian mantle. Depending on the assumed density contrast between primitive mantle and residual mantle (mantle material depleted in crustal components during the melting process), two to four distinct reservoirs may form early in the planetary evolution. Left panels show the concentration of radioactive elements in the crust and mantle for the two scenarios after 4.5 Ga. Reproduced from Plesa AC and Breuer D (2014) Partial melting in one-plate planets: Implications for thermo-chemical and atmospheric evolution. *Planetary and Space Science*, special issue 'Planetary evolution and life', 98: 50–65, http://dx.doi.org/10.1016/j.pss.2013.10.007.

and during the aerobraking phase and science-phasing orbit phase, it was possible to get magnetic measurements at periapses ranging from 85 to 170 km above Mars' surface. Connerney et al. (1999) (see also Acuña et al., 2001; Purucker et al., 2000) had discovered strongly magnetized regions in the crust, closely associated with the ancient, cratered terrain of the highlands in the southern hemisphere. There has been a great interest in the meaning of the spatial pattern of magnetization (Arkani-Hamed, 2004; Jurdy and Stefanick, 2004), including possible lineations (these crustal magnetizations even suggest reversals) that were believed to suggest an analogy to plate tectonics (Connerney et al., 1999, 2005), but other models favor thermal remanence acquired during the time these ancient rocks cooled in conjunction with dike intrusions (Nimmo, 2000). The inferred crustal magnetizations, which are up to $\sim$10–30 A m^{-1}, are an order of magnitude stronger than those of the Earth's continents (Arkani-Hamed and Dyment, 1996; Toft and Arkani-Hamed, 1992) and comparable in magnitude with the remanent magnetization of fresh extrusive basalt near the oceanic ridge axes of Earth (Bleil and Petersen, 1983). These magnetic anomalies indicate the existence of a strong ancient intrinsic Martian magnetic moment corresponding to a magnetic field strength of 0.1–10 times that of the Earth's at present (Mitchell et al., 1999, 2001; Ness et al., 1999).

There are several ways in which the Martian crust could have been magnetized. One of the most probable ways is thermoremanent magnetization (TRM). With TRM, the magnetization is produced when rock cools below a critical temperature (Curie temperature) in the presence of a magnetizing field. This is an effective mode for producing an intense remnant field. The magnitude of magnetization produced in the rock depends on the strength of the internal field, the mineralogy, and the magnetic microstructure. There is a trade-off between the concentration of magnetic carriers and the strength of the magnetic field: The lower the imposed magnetic field, the more magnetic carriers are required to explain an observed magnetization. Assuming an early Martian magnetic field similar to the present-day Earth field, the concentration of magnetic carriers might be comparable to the amount in extrusive basalt. Bearing in mind the possible reduction of the remnant magnetization through viscous decay and chemical alterations during this very long geologic time (the magnetization of the oceanic basalt decreases by a factor of 4–5 during the first 20 My, largely by chemical alterations (Bleil and Petersen, 1983)), the initial magnetization of the Martian source bodies must have been even stronger. However, there is ample evidence that the FeO content of the Martian mantle is about twice that of the Earth's mantle (Sanloup et al., 1999; Sohl and Spohn, 1997). Whether this high concentration of FeO translates to a high concentration of magnetic minerals depends on the oxidation state of the Martian mantle and lower parts of the crust. However, for a high concentration of magnetic minerals, the early magnetic field could have been smaller than that of the present Earth.

The timing of the dynamo places strong constraints on the thermal evolution of the planet. It has been proposed on the basis of the surface distribution of the magnetic anomalies that this magnetization event predated the formation of the Hellas and Argyre basins (Connerney et al., 1999; Lillis et al., 2013), roughly at the end of the Early Noachian (Head et al., 2001). The existence of an early magnetic field has been supported by the ancient Martian meteorite ALH84001, which shows a magnetization that was acquired at 4 Ga or even earlier (Weiss et al., 2002). However, the lack of an appreciable magnetization in two ancient areas on the highlands, the South Province and the Tempe Terra, further suggests that no strong dynamo existed in the first $\sim$100 My of Mars' history (Arkani-Hamed and Boutin, 2012). It has been alternatively suggested that the dynamo turned on after the giant impact craters formed, implying that the remnant magnetization of the crust was acquired later in Martian history (Schubert et al., 2000). This hypothesis is supported by the SNC meteorites with formation ages of 1.3 Ga to 180 Ma, which show a remnant magnetization consistent with a weak ancient surface field of a few hundred to a few thousand nanoteslas (Collinson, 1997; Rochette et al., 2001). This weak surface field may suggest a late onset or reactivation of a core dynamo that would have stopped operating before the present. However, it is also possible that the young Martian meteorites were magnetized by the older and strongly magnetized crustal rock (see Connerney et al., 2003, for a review).

The evidence of early dynamo action is supported by most thermal evolution models that consider the magnetic field history (e.g., Breuer and Spohn, 2003, 2006; Hauck and Phillips, 2002; Michel and Forni, 2011; Schubert and Spohn, 1990; Spohn et al., 1998; Williams and Nimmo, 2004). The models assume a thermally driven dynamo during the first few hundred million years as the consequence of the early rapid cooling if the core is superheated with respect to the mantle as a consequence of the rapid core formation process (**Figure 15**) – although these models are at variance with a later dynamo start after the first 100 My (Arkani-Hamed and Boutin, 2012). In the subsequent evolution, the core remains liquid with a heat flow below the critical value and, thus, no dynamo action

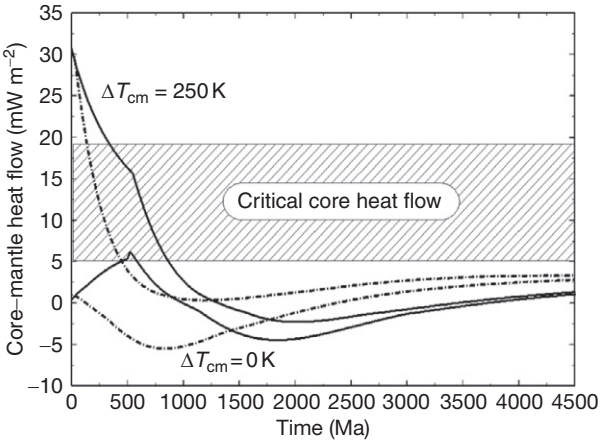


Figure 15 Core–mantle heat flow as a function of time for models with early plate tectonics (solid line) and models with stagnant lid convection throughout the entire evolution (dash-dotted line) with initial temperature differences across the CMB of ΔT_{cm} = 0 and 250 K. Reproduced from Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research: Planets* 108(E7): 5072. http://dx.doi.org/10.1029/20002JE001999.

until today. A totally molten core for the entire evolution is possible if the core contains at least ~5 wt% sulfur (Williams and Nimmo, 2004), although the more recent melting data of Fe–FeS by Stewart et al. (2007) suggest a sulfur threshold of at least 10% or even higher depending on the efficiency of cooling. A core sulfur abundance of 14.2 wt% is suggested for Mars from geochemical analysis of the SNC meteorites (Dreibus and Wänke, 1985; McSween, 1985). Assuming a sulfur content in the core below 10 wt%, inner core growth and consequently a chemical dynamo would be possible. Such a scenario, however, is at variance with the assumption of a brief, very early dynamo. If core solidification did take place on Mars, it likely would have generated a very long-lived dynamo (Schubert et al., 1992) unless the iron snow is the prevalent freezing mechanism in the Martian core as suggested by Stewart et al. (2007). For this freezing mechanism, the magnetic field generation and its possible duration are not well understood.

An alternative scenario assumes a change in the heat transport mechanism of Mars as suggested by the model of early plate tectonics (Nimmo and Stevenson, 2000; Stevenson, 2001). During the phase of plate tectonics, the interior cools efficiently and not only a thermal dynamo but also a chemical dynamo is possible if the core cooled rapidly below its melting temperature. As soon as this regime is followed by stagnant lid convection, the mantle heats up and dynamo action stops until today (**Figure 15**). In the case of a thermal dynamo, the heat flow decreases below the critical value. In the case of a chemical dynamo, the end of the plate tectonic regime marks the time of the coldest interior. Thus, in the subsequent evolution, the core temperatures remain above the melting temperature and inner core growth and chemical dynamo action stop.

Impacts have been also identified to play a major role in the magnetic field generation. Some models predict that large impacts forming basins over 2500 km in diameter cease the dynamo action due to the heating of the mantle or the outermost core (Arkani-Hamed and Olson, 2010; Roberts et al., 2009). Such a thermal state suppresses core cooling that is unfavorable for core convection and could have led to the cessation of dynamo activity. Alternatively, it has been suggested that the thermal anomaly induced by a large impact and the formation of a hot molten iron layer from the impactor's core at the CMB can favor dynamo generation (Reese and Solomatov, 2010). Which scenario is most likely depends mainly on the dynamics in the core during merging of an iron diapir and a preexisting core – a process that is not well understood (Monteux et al., 2013).

The magnetic field data by MGS revealed that the Martian crust is strongly magnetized in the southern hemisphere while the northern hemisphere only shows weak signals (Langlais et al., 2004). Two possible reasons have been suggested for this dichotomy: A once more or less homogeneous magnetization may have been destroyed in the northern hemisphere by, for example, hydrothermal alteration (Solomon et al., 2005) or impacts (e.g., Mohit and Arkani-Hamed, 2004; Rochette et al., 2003). An alternative theory suggests that the dynamo itself produced a hemispheric field due to lateral heat flow variations at the CMB (Amit et al., 2011; Stanley et al., 2008). This hemispheric dynamo would produce strong magnetic fields in the southern hemisphere in case the heat flow out of the core below the southern hemisphere was stronger than below the northern hemisphere. A recent numerical model by Dietrich and Wicht (2013) shows, however, that these hemispheric dynamos reverse on timescales of about 10 ky – too fast to allow for a unidirectional magnetization of a thick crustal layer required to explain the strong magnetization in the southern hemisphere.

10.08.7 Io

10.08.7.1 Tidal Heating and Thermal Evolution

Io is unique in the solar system in that its primary source of internal heat is not radioactive decay but tidal dissipation (Peale et al., 1979). Io is the Galilean satellite nearest to Jupiter and it is the most volcanically active body in the solar system. Unlike most bodies in the solar system, which would require in situ methods to determine the heat flow, Io's surface heat flow can be estimated from remote observations of its thermal emission. From a global inventory of Io's thermal radiation from a variety of sources, McEwen et al. (2004) obtained a global heat flow estimate of $2.1 \pm 0.7\ \mathrm{W\,m^{-2}}$, while Veeder et al. (2004) gave $3 \pm 1\mathrm{W\,m^{-2}}$. The true value most likely falls in the range from 1.5 to 4 $\mathrm{W\,m^{-2}}$ (60–160 TW total). The heat production rate implied by the surface heat flow far exceeds that that can plausibly be produced by radioactive decay; thus, a different internal heat source must be active. Io's location so close to Jupiter causes it to be subject to intense tidal forces. The constantly varying tidal potential resulting from Io's eccentric orbit results in more than 10 m radial deformation of Io's solid surface every orbit (1.8 days). Viscoelastic dissipation of heat in the interior of Io due to the repeated tidal deformation is responsible for Io's large internal heat source. The tremendous heat flow measured at the surface of Io is the result of silicate volcanism, with typical inferred eruption temperatures of 1200–1400 K and extreme temperatures over 1800 K (McEwen et al., 1998). It is likely that the sulfur-based volcanism observed by Voyager is also driven by silicate magma.

Since the distribution of tidal heat within Io is strongly dependent on the variation of viscosity (Segatz et al., 1988), it has been proposed to use the distribution of volcanoes to infer the viscosity structure. The process of getting heat to the surface, however, is influenced by mantle convection, which will act to smooth out variations, and melt transport, which will be sensitive to lithospheric structures. Tackley et al. (2001) demonstrated that, even at high Rayleigh number, the spatial pattern of heating within the mantle should be reflected at the surface, and they favored an asthenospheric source for the heating. Hamilton et al. (2013) analyzed the distribution of volcanic centers based on clustering statistics and found that an asthenospheric model was preferred; however, a consistent 30–60° eastward offset of the observed pattern indicated that lithospheric structure, perhaps offset by nonsynchronous rotation, played an important role.

Despite the intensity of the volcanic activity on Io and the inferred high temperatures of its interior, Io supports some of the highest and steepest topography of any planetary body. Mountains up to 17 km high dot the surface (Jaeger et al., 2003; Schenk et al., 2001), and oddly, these are not volcanoes.

Instead, volcanic centers are generally depressions, while the mountains appear to be fault-bounded tectonic uplifts. The presence of such large topography seems incompatible with a lithosphere that must allow the transport of $\sim 2.5\ \mathrm{W\ m^{-2}}$ of heat.

A solution to this problem was found by O'Reilly and Davies (1981), who proposed a heat-pipe mechanism for heat transport through Io's lithosphere accomplished by melt transport through fissures. The most important implication of this model is that, as successive eruptions spread across the surface and cool, they bury the flows of previous eruptions, resulting in an advective transport of cold material from the surface downward. In order to match the observed heat flow, Io must be resurfaced globally by silicate magma to a depth of 1–2 cm every year. This is then the rate at which material is advected downward within the lithosphere. The equation for the temperature T in the lithosphere is then

$$\frac{k}{\rho c_\mathrm{p}}\frac{\mathrm{d}^2T}{\mathrm{d}z^2} = v\frac{\mathrm{d}T}{\mathrm{d}z} - \frac{H}{\rho c_\mathrm{p}} \qquad [23]$$

where k is the thermal conductivity, ρ is the density, c_p is the specific heat, z is the depth from the surface, v is the downward velocity (resurfacing rate), and H is the volumetric heat production. The temperature must match the surface temperature T_s at the surface and the melting temperature T_m at the base of the lithosphere. The solution in the case of no heat production is

$$T(z) = T_\mathrm{s}\frac{(T_\mathrm{m} - T_\mathrm{s})\left(\exp^{\lambda\xi} - 1\right)}{\exp^{\lambda} - 1} \qquad [24]$$

where $\xi = z/D$ is the depth normalized by the thickness of the lithosphere D and the dimensionless parameter $\lambda = Dv\rho c_\mathrm{p}/k$ is the advective velocity normalized by the conductive velocity scale. This can be related to the heat flux carried by melting F as follows:

$$\lambda = \frac{Dc_\mathrm{p}F}{k\left[L_\mathrm{f} + c_\mathrm{p}(T_\mathrm{m} - T_\mathrm{s})\right]} \qquad [25]$$

where L_f is the latent heat of fusion of the silicate rocks. For a heat flux of $2.5\ \mathrm{W\ m^{-2}}$ and a 30 km thick lithosphere, λ is about 10. Solutions for different values of λ are shown in Figure 16. For λ of 3, more than 80% of the thickness of the lithosphere is below 900 K, and for λ of 10, more than 95% of the lithosphere is cold enough to sustain elastic stresses for very long periods of time. This is how Io's lithosphere can support huge mountains while at the same time allowing a heat flux of $2.5\ \mathrm{W\ m^{-2}}$ to pass through.

What sort of melt fractions does the heat-pipe mechanism imply for the asthenosphere? It is straightforward to calculate the melt segregation velocity due to Darcy flow driven by the buoyancy of the melt (e.g., Scott and Stevenson, 1986):

$$\Phi v = \frac{k_\Phi g \Delta\rho}{\eta_\mathrm{m}} \qquad [26]$$

where Φ is the melt volume fraction (porosity), g is the gravitational acceleration, $\Delta\rho$ is the difference between the solid and melt densities ($\sim 500\ \mathrm{kg\ m^{-3}}$), η_m is the melt viscosity (~ 1000 Pa s), and k_Φ is the permeability, which is related to the porosity by a function of the form

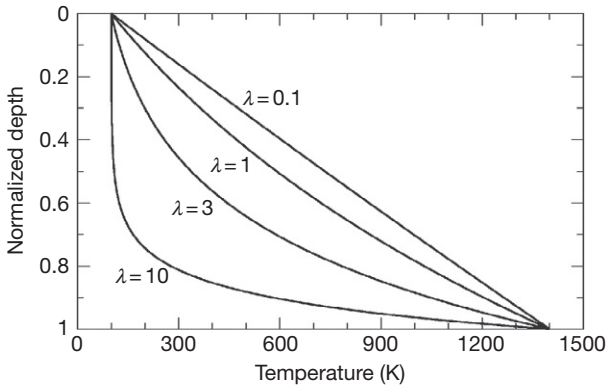


Figure 16 Temperature as a function of normalized depth for different values of λ; the normalized advective velocity defined in the text.

$$k_\Phi = \frac{b^2\Phi^n}{\tau} \qquad [27]$$

where b is a typical grain size (~ 1 cm) and n and τ are constants that are functions of the geometry of the melt. The dependence of k_Φ on grain size is overly simplified in this model, since real systems may have broad grain size distributions. Using the $1\ \mathrm{cm\ year^{-1}}$ resurfacing velocity (which is actually the melt flux Φv) and inserting experimentally determined values for the constants n (3) and τ (200) (Liang et al., 2001; Wark and Watson, 1998) result in an estimate for Φ between 10% and 20% (Moore, 2001). A global subsurface magma layer with a thickness of over 50 km and a rock melt fraction of 20% or more has been suggested by the magnetometer data of the Galileo spacecraft (Khurana et al., 2011). Melt fractions exceeding this will transport too much heat (the heat transported goes as k_Φ), thus cooling the asthenosphere and bringing the melt fraction back to the equilibrium value.

A rough estimate for the thickness of the asthenosphere may be obtained by extrapolating along the adiabat from the temperature at the top of the melt zone to the solidus:

$$d = \frac{\Delta T_\mathrm{m}}{(\mathrm{d}T/\mathrm{d}z)_\mathrm{m} - (\mathrm{d}T/\mathrm{d}z)_\mathrm{ad}} \qquad [28]$$

where $\Delta T_\mathrm{m} = \Phi L_\mathrm{f}/c_\mathrm{p}$ is the temperature excess required to reach a melt fraction Φ and the denominator is the difference between the slopes of the solidus ($\sim 0.8\ \mathrm{K\ km^{-1}}$) and the adiabat ($\sim 0.1\ \mathrm{K\ km^{-1}}$). Using the melt fractions estimated earlier in the text, the asthenosphere is 60–120 km thick. Below this is a mantle that is solid, but quite close to the solidus. Though we lack direct evidence, the core is most likely liquid, since even the pure iron solidus is below the silicate solidus. This simple analysis of melt segregation ignores the role of convection in heat transport within Io. Despite its broad applicability to planetary heat transport, however, heat transport in partially molten, convecting systems has not received thorough quantitative treatment. Unlike subsolidus convection, for which detailed parameterizations exist that incorporate many aspects of the physics of planetary materials (particularly the rheology of rocks), partially molten convection has been treated in an ad hoc fashion by tacking a melt segregation analysis such as that presented previously in the text onto a subsolidus convection parameterization.

This approach ignores the strong influence of melting and melt segregation on the convective flow itself, which is particularly important when the rheology of the convecting layer influences the heat production as it does in a tidally heated body such as Io.

10.08.7.2 Thermal and Orbital Evolution

The tidal heat source not only is capable of much greater heating rates but also evolves with the orbit of Io, since it depends on both the distance of Io from Jupiter (the semimajor axis) and the eccentricity of Io's orbit. Tidal heating is also very sensitive to the rheology of Io's interior. The heat source is therefore coupled to both the orbit and the interior temperature. This results in a very different thermal evolution for Io than for other bodies of similar size (e.g., the Moon).

The tidal heating of Io's interior is a result of the resonant orbital interactions between Io, Europa, and Ganymede, which have orbital periods that are in the ratio τ_I: τ_E: $\tau_G = 1:2:4$. The orbits of all three satellites have evolved together into what is called the Laplace resonance, where the conjunctions of each satellite with its nearest outward neighbor occur when the inner satellite is at perijove (nearest to Jupiter) and the outer satellite is at apojove (furthest from Jupiter), thus maximizing their separation. These repetitive alignments cause the orbits to remain elliptical, even though dissipation of tidal energy in the satellites should tend to circularize their orbits (while dissipation in Jupiter tends to make them more eccentric). In addition to the effects of dissipation on eccentricity, tides raised on the satellite tend to reduce the semimajor axis of the orbit, that is, the satellite tends to spiral inward. Thus, Io's orbital evolution is tightly coupled to the thermal evolution of its interior. Conversely, tides raised on Jupiter by the satellite cause the satellite to spiral outward (like the Moon). Thus, Io's actual rate of orbital migration represents a balance between dissipation in Io and dissipation in Jupiter. Indeed, this orbital migration is thought to be the process by which the resonant configuration is assembled (Yoder, 1979).

The coupled thermal–orbital evolution of Io (and Europa and Ganymede) is a complex dynamic system with a wide possible range of behaviors. The interior temperature is set by the balance between tidal heating and heat transport (convective or magmatic). Orbital evolution is driven by dissipation in Io. Fischer and Spohn (1990) studied the coupled system including a simplified orbital evolution (linearized about the Laplace resonance) and convective heat transport and recognized that there are multiple equilibriums for the thermal state of Io, as illustrated in **Figure 17**, a cartoon depicting the possible thermal equilibriums in a tidally heated body. Radioactivity (solid line) produces the same heating regardless of temperature, while tidal heat production (solid curve) depends strongly on temperature, increasing with temperature until the Maxwell time (viscosity over shear modulus) of the material approaches the tidal forcing period. At a critical temperature (T_c) slightly above the solidus, the shear modulus begins to decrease rapidly (Berckhemer et al., 1982), resulting in a drop in tidal heating. At the breakdown temperature T_b, the solid matrix loses coherence, and the material becomes dominated by the mechanical properties of the liquid and the tidal heating is greatly reduced. Convective heat transport (dashed curve)

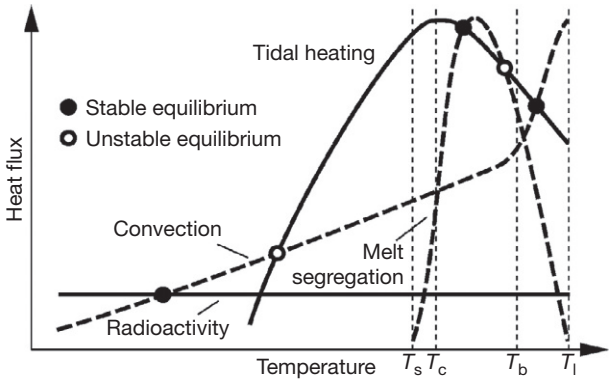


Figure 17 Cartoon depicting the possible thermal equilibriums (circles) in a tidally heated body. Heat sources (radioactivity and tidal heating) are shown as solid lines, and heat transport mechanisms (convection and melt segregation) are shown as dashed lines. Also marked schematically are the solidus T_s, critical T_c, breakdown T_b, and liquidus T_l temperatures.

increases with temperature, as the viscosity decreases, becoming extremely efficient as the liquidus temperature T_l is approached. Melt segregation (as described in the preceding text) can only occur between the solidus and liquidus and is not efficient above the breakdown temperature (since there is no longer any solid matrix for the melt to segregate from). Depending on the details of the heat production and transport processes, not all of these equilibriums may be realized, so we use a qualitative diagram to illustrate the possible range of behaviors. Fischer and Spohn (1990) identified the highest and lowest temperature equilibriums (the points where the solid tidal heat and radioactivity curves intersect the dashed convection curve) and noted that at times of low eccentricity, the high-temperature equilibrium may cease to exist. This is because low eccentricity shifts the tidal heat production curve downward, and it may fail to intersect the convective heat transport curve at all. Such episodes are followed by rapid cooling, as Io evolves toward the more common radioactivity-driven convective equilibrium. Moore (2003) investigated the tidal-convective equilibrium (for the current eccentricity, independent of the orbital evolution) and concluded that the high-temperature equilibrium, while present, cannot explain the observed heat flow; thus, a different heat transport process (melt segregation) was necessary.

Melt segregation can introduce additional equilibriums as shown in **Figure 17**. These equilibriums are necessarily between the solidus and liquidus temperatures (T_s and T_l). The simplified calculation given in **Section 10.08.7.1** required melt fractions of 10–20% to achieve equilibrium. This equilibrium is stable but may disappear at times of low eccentricity. Since this equilibrium is reached at temperatures lower than the convective equilibrium (this is a consequence of the high viscosity of rocks, even when partially molten), it is likely that Io never reaches the high-temperature convective equilibrium. Io's volcanic activity and high heat flow strongly suggest that Io is in (or near) an equilibrium state with melt segregation balancing tidal heat production. It is not known whether our roughly three decades of heat flow observations are representative of the long-term average, but if 10^{14} W is the rate of

energy dissipation in Io, then if we know the rate of dissipation in Jupiter, we can predict the rate at which Io's orbit is evolving. Unfortunately, the mechanisms by which tidal motions are dissipated in a deep atmosphere such as Jupiter's are not understood at present. We are therefore left with trying to measure the orbital evolution to estimate dissipation in Jupiter. A recent determination of the orbital accelerations of the Galilean satellites (Lainey et al., 2009) yields an orbital acceleration ($\dot{n}/n$) of $0.14 \pm 0.01 \times 10^{-10}$, a value consistent with the present heating rate. We can put certain bounds on the total outward migration of Io's orbit (assuming Europa and Ganymede have moved in concert) by assuming Io formed as close to Jupiter as possible (2.4 R_J) and evolved to its current orbital distance over the age of the solar system. This upper bound on the migration rate, however, is still much too low to account for Io's current dissipation rate, which depends on the maintenance of a high eccentricity (and thus high dissipation in Jupiter) (Yoder and Peale, 1981). We are therefore forced to conclude that Io's orbital evolution has not been steady over the age of the solar system or that the current heat flow does not represent a long-term average of the tidal dissipation. Either case is possible, since we do not know if the Laplace resonance is primordial or if it was assembled more recently by orbital migration. Furthermore, Io's surface heat flow represents dissipation integrated over some unknown transport time. However, if most of the heat is carried by melt segregation, as seems to be the case (in the preceding text and Moore (2003)), the transport time is short (tens of years), and the heat flow does reflect recent tidal dissipation (Lainey et al., 2009). A longer transport time regulated by convective motions would allow for the possibility of oscillatory states in which the heat flow and dissipation are out of phase (Fischer and Spohn, 1990; Hussmann and Spohn, 2004).

10.08.8 Summary

This chapter has reviewed the study of the interior structure, chemistry, and dynamics of terrestrial planetary bodies: Mercury, the Moon, Venus, Mars, and Io. There are many processes that connect these aspects of planetary interiors, many of which have been discussed here. Fundamentally, the dynamics of terrestrial bodies are driven by the transport of heat from the interior to the surface. Heat generation and transport are the processes that connect the structure and chemistry of a planetary body with its dynamic behavior, for example, volcanism, tectonism, and magnetism. The physical and chemical properties of planetary materials influence the mechanics of the planetary heat engine driving dynamics, while the dynamics in turn alter the physical and chemical properties of the planet's interior.

The density structure of each of the terrestrial bodies is roughly similar, with silicate mantles surrounding metallic cores, but the geometries vary significantly, from shallow mantles (Mercury) to very small cores (the Moon). The aspect ratio of the mantle can have a strong influence on the planform of convective motions in weakly convecting systems. The silicate mantles are similarly differentiated into crustal and mantle portions, with considerable variation in the thickness of the crust. The process of crustal formation, while universal, has not been identical on each of the terrestrial bodies. Both primordial (the Moon and Mercury) and evolved (Venus and Io) crusts exist today. This likely reflects the importance of crustal recycling and continuous crustal formation on the more active bodies. Crustal differentiation may have several different effects on the planetary heat engine. First, as a form of heat transport, melt segregation can be extremely efficient as seen at Io. For less active bodies, the crust can act both as an insulator for the mantle and as a sink for heat-producing elements. These effects have opposite influences on the temperature of the underlying mantle; the former tends to raise the temperature while the latter depresses it. Crustal material that is unevenly distributed may influence the temperature distribution in the upper mantle and the planform of convection, perhaps explaining the distribution of volcanism on Mars. Finally, melt segregation from the mantle may also be associated with devolatilization of the interior, which influences the rheology of the remaining mantle material.

The thermal evolution of the terrestrial planetary bodies through the history of the solar system has been studied by using parameterizations of heat transport through their mantles, but recently, modelers started to also use 2-D and 3-D convection models. Many physical processes have been incorporated in these models, including complex rheology (with temperature, pressure, stress, and composition dependence), partial melting, crustal formation, core heat flux, and inner core growth. These models are compared with observations of crustal and lithospheric properties, magnetic fields, tectonism, and volcanic activity. Significant uncertainties in the properties of the interiors of the terrestrial planets remain, so only general conclusions can be reached.

Mercury's magnetic field, if produced by dynamo action in the core, provides a very strong constraint on the heat transport through the mantle. If freezing of an inner core is required for dynamo action, this also places a constraint on the sulfur content of the core. Mercury's surface also records a few kilometers of radial contraction over its history in the form of lobate scarps. In order to generate a present-day magnetic field without completely freezing the core, convection in Mercury's mantle must be very weak or possibly have ceased altogether. A stiff rheology and a thin isolating crust seem to be required.

The situation on Venus is less well constrained, in part because much of the evidence of Venus' early history has been erased from the surface. Venus has been recently active (within the last 10^8 years and perhaps as recently as 10^5 years), but the cratering record is permissive of a wide variety of resurfacing scenarios. The strong correlation of gravity and topography at low harmonic degrees indicates that there are large-scale density anomalies in the mantle, likely the result of mantle convection. Whether Venus ever experienced a plate tectonic phase (or even several) is uncertain, but there is little present evidence for active plate tectonics. The observation that Venus lacks a magnetic field may be related to the relatively low heat transport efficiency of a single-plate planet with a rigid lid.

Unlike the more active bodies, which largely erase the influence of their initial conditions, the present state of the Moon may be strongly reflective of its formation. The thick lunar crust appears to be a primordial differentiate from a magma ocean that formed during the accretion of the Moon

from a hot debris disk orbiting the Earth after a giant impact. The differentiation of the magma ocean may have led to an initially unstable density stratification that drove an early episode of convection, but the segregation of heat-producing elements out of the mantle may result in only weak convective motions and fairly little internal cooling since then. The lunar core is small (< 500 km), and its composition is unknown. Certain lunar rocks have been magnetized, but whether this is evidence of an internally generated magnetic field is not certain. If there was an early lunar dynamo, it seems to have gradually decayed with time, beginning at about 4 Ga and ceasing at about 3 Ga. This perhaps reflects an early period of vigorous convection, but some mechanism is required to prevent dynamo onset earlier than 4 Ga unless the earliest remanent magnetizations are not measurable or existent although an early dynamo action was active.

Our knowledge of the interior and the evolution of Mars has been vastly improved by recent satellite and lander missions. The size of the Martian core is now better constrained, as is the average crustal thickness. Due to the ambiguity in these parameters as we do not know the composition and density of the core, mantle, and crust, only future seismic experiments may further improve our knowledge. Like the Moon, Mars has ancient, heavily cratered highlands, but unlike the Moon, the less cratered lowlands are not clearly impact basins. The ancient highlands, however, are regions of thickened crust, suggesting a primordial origin, though the mechanism and exact timing of the dichotomy formation are still debated. Geochemical evidence from SNC meteorites suggests that little mixing has gone on since early in Mars' history and thus that the bulk of the Martian crust was formed very soon after accretion. Furthermore, there are more and more data and models suggesting a wet interior – with a volatile content at least influencing the rheology if not even the petrology – that degassed with time due to crustal formation. Mars has been volcanically active into the recent past, and there are a wide range of tectonic features mostly related to volcanic construction. The discovery of strong magnetic anomalies associated with magnetized crust has led to new models of the early thermal history of Mars that include an episode of plate tectonics or an initially superheated core due to its formation to provide the required core cooling. Since that epoch, the activity of the Martian interior has been gradually declining. The crustal dichotomy provides a spatially variable boundary condition to the convecting mantle and may provide a means for keeping some partial melt and for focusing it at Tharsis.

Although it is an outer planet satellite, Io is essentially a terrestrial body, with a silicate mantle and a metallic core. Io is unique, however, in that its internal heat source is tidal dissipation rather than radioactive decay. This ties the heat production in the interior to the internal physical state through the rheology. Io is also unique in that the dependence of tidal heating on the material properties is such that the internal temperature is kept above the solidus. This leads to massive amounts of volcanic activity and makes Io the only terrestrial body to transport heat primarily by melt segregation. Io today may provide us an opportunity to study processes that were in effect on the other terrestrial bodies only briefly in their earliest histories. The constant resurfacing by volcanic eruption has erased all evidence of impacts on Io's surface and produces strong compressional stresses as the surface subsides beneath subsequent flows. These stresses lead to large fault-bounded massifs, some of the tallest in the solar system. Ionian volcanoes, however, are broad, low-relief structures, perhaps because the abundant sulfur deposits on the surface and in the crust are easily weakened by the locally high heat flow.

The study of terrestrial planet interiors has advanced significantly in the last decades, driven by improvements in modeling, increasing computing power, and new observations of Mercury, Venus, Mars, the Moon, and Io. The coming decades will see additional missions to Mercury, the Moon, Mars, and the outer solar system that will expand our knowledge of terrestrial bodies. Extrasolar terrestrial planets allow us to put the formation of our own planetary system in a broader context. Until today, four exoplanets, that is, CoRoT-7b (Bruntt et al., 2010; Hatzes et al., 2011; Léger et al., 2009), Kepler-10b (Batalha et al., 2011), Kepler-36b (Carter et al., 2012), and Kepler-78b (Howard et al., 2013; Pepe et al., 2013), have been detected, from which we know mass and radius and that can be characterized as terrestrial. Our understanding of the properties and behavior of planetary materials, especially at high temperature and pressure, will continue to grow through experimentation and observation. Looking further ahead, there will come a day when terrestrial planetary bodies will be directly explored by human beings, and a new era of terrestrial planet geology will dawn. Though that day may be far in the future, that achievement will build upon the research summarized in this chapter, extending and refining our understanding of the terrestrial planets.

Acknowledgments

We thank Stephen Hauck and Gregor Golabek for their helpful reviews that improved the manuscript.

References

Acuña MH, Connerney JEP, Ness NF, et al. (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars Global Surveyor Mission. *Science* 279: 1676–1680.

Acuña MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.

Acuña MH, Connerney JEP, Wasilewski P, et al. (2001) Magnetic field of Mars: Summary of results from the aerobraking and mapping orbits. *Journal of Geophysical Research* 106(E10): 23403–23418.

Albers M and Christensen UR (1996) The excess temperature of plumes rising from the core–mantle boundary. *Geophysical Research Letters* 23: 3567–3570.

Alley KM and Parmentier EM (1998) The development of mantle convection in an initially stable chemical stratification; Application to a model of lunar mare basalt genesis. *Physics of the Earth and Planetary Interiors* 108: 15–32.

Amit H, Christensen UR, and Langlais B (2011) The influence of degree-1 mantle heterogeneity on the past dynamo of Mars. *Physics of the Earth and Planetary Interiors* 189: 63–79.

Anderson DL (1987) A seismic equation of state II: Shear properties and thermodynamics of the lower mantle. *Physics of the Earth and Planetary Interiors* 45: 307–323.

Anderson JD (1998) Gravitational fields and the interior structure of Icy Satellites. American Physical Society, Joint Fall Meeting of the Texas Section of the APS, Abstract #2P.03.

Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The Mass Gravity Field and Ephemeris of Mercury. *Icarus* 71(3): 337–349.

Anderson BJ, Johnson CL, Korth H, et al. (2011) The global magnetic field of Mercury from MESSENGER orbital observations. *Science* 333(6051): 1859–1862.

Anderson JD, Jurgens RF, Lau EL, Slade MA III, and Schubert G (1996) Shape and orientation of Mercury from radar ranging data. *Icarus* 124(2): 690–697.

Anderson F and Smrekar S (2006) Global mapping of crustal and lithospheric thickness on Venus. *Journal of Geophysical Research* 111(E8). http://dx.doi.org/10.1029/2004JE002395.

Andrews-Hanna JC, Asmar SW, Head JW, et al. (2013) Ancient igneous intrusions and early expansion of the Moon revealed by GRAIL gravity radiometry. *Science* 339: 675–678. http://dx.doi.org/10.1126/science.1231753.

Andrews-Hanna JC, Phillips RJ, and Zuber MT (2007) Meridiani Planum and the global hydrology of Mars. *Nature* 446(7132): 163–166.

Andrews-Hanna JC, Zuber MT, and Banerdt WB (2008) The Borealis basin and the origin of the martian crustal dichotomy. *Nature* 453(7199): 1212–1215.

Arkani-Hamed J (2004) A coherent model of the crustal magnetic field of Mars. *Journal of Geophysical Research* 109(E9), CiteID E09005.

Arkani-Hamed J and Boutin D (2012) Is the primordial crust of Mars magnetized? *Icarus* 221(1): 192–207.

Arkani-Hamed J and Dyment J (1996) Magnetic potential and magnetization contrasts of Earth's lithosphere. *Journal of Geophysical Research* 101(B5): 11401–11426.

Arkani-Hamed J and Olson P (2010) Giant impact stratification of the Martian core. *Geophysical Research Letters* 370(L02): 201.

Arkani-Hamed J, Schaber GG, and Strom RG (1993) Constraints on the thermal evolution of Venus inferred from Magellan data. *Journal of Geophysical Research* 98: 5309–5315.

Armann M and Tackley PJ (2012) Simulating the thermo-chemical magmatic and tectonic evolution of Venus' mantle and lithosphere 1. Two-dimensional models. *Journal of Geophysical Research* 117, E12003. http://dx.doi.org/10.1029/2012JE004231.

Arndt NT and Goldstein SL (1989) An open boundary between lower continental crust and mantle: Its role in crust formation and crustal recycling. *Tectonophysics* 161(3): 201–212.

Babeyko AY and Zharkov VN (2000) Martian crust: A modeling approach. *Physics of the Earth and Planetary Interiors* 117(1): 421–435.

Baker J, Bizzarro M, Wittig N, Connelly J, and Haack H (2005) Early planetesimal melting from an age of 4.5662 Gyr for differentiated meteorites. *Nature* 436: 1127–1131.

Balta JB and McSween HY (2013) Water and the composition of martian magmas. *Geology* 41(10): 1115–1118. http://dx.doi.org/10.1130/G34714.1.

Banerdt WB and Golombek MP (2000) Tectonics of the Tharsis region of Mars: Insights from MGS topography and gravity. In: *31st Annual Lunar and Planetary Science Conference*, Houston, TX, 13–17 March 2000, Abstract No. 2038.

Baratoux D, Toplis MJ, Monnereau M, and Gasnault O (2011) Thermal history of Mars inferred from orbital geochemistry of volcanic provinces. *Nature* 472: 338–341. http://dx.doi.org/10.1038/nature09903.

Basaltic Volcanism Study Project (1981) *Basaltic Volcanism on the Terrestrial Planets*. New York: Pergamon Press, 1286 pp.

Batalha NM, Borucki WJ, Bryson ST, et al. (2011) Kepler's first rocky planet: Kepler-10b. *Astrophysical Journal* 729(1): 27.

Belleguic V, Lognonne P, and Wieczorek M (2005) Constraints on the martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005. http://dx.doi.org/10.1029/2005JE002437.

Berckhemer H, Kampfmann W, Aulbach E, and Schmeling H (1982) Shear modulus and Q of forsterite and dunite near partial melting from forced-oscillation experiments. *Physics of the Earth and Planetary Interiors* 29: 30–41.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Physics of the Earth and Planetary Interiors* 205: 107–121.

Bercovici D, Schubert G, and Tackley PJ (1993) On the penetration of the 660 km phase change Mantle downflows. *Geophysical Research Letters* 20: 2599–2602.

Berman DC and Hartmann WK (2002) Recent fluvial, volcanic, and tectonic activity on the Cerberus Plains of Mars. *Icarus* 159: 1–17.

Binder AB (1986) The binary fission origin of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 499–515. Houston, TX: Lunar and Planetary Institute.

Binder AB and Lange MA (1980) On the thermal history, thermal state, and related tectonism of a moon of fission origin. *Journal of Geophysical Research* 85: 3194–3208.

Bizzarro M, Baker JA, and Haack H (2004) Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* 431: 275.

Bjonnes EE, Hansen VL, Jansen B, and Swenson JB (2012) Equilibrium resurfacing of Venus: Results from new Monte Carlo modeling and implications for Venus surface histories. *Icarus* 217: 451–461.

Bland MT, Showman AP, and Tobie G (2009) The orbital–thermal evolution and global expansion of Ganymede. *Icarus* 200(1): 207–221.

Bleil U and Petersen N (1983) Variations in magnetization intensity and low-temperature titanomagnetite oxidation of ocean floor basalts. *Nature* 301: 384–388.

Bond TM and Warner MR (2006) Dating Venus: Statistical models of magmatic activity and impact cratering. In: *37th Annual Lunar and Planetary Science Conference*, Abstract No. 1957.

Borg LE, Nyquist LE, Wiesmann H, and Shih C-Y (1997) Constraints on Martian differentiation processes from Rb–Sr and Sm–Nd isotopic analyses of the basaltic shergottite QUE94201. *Geochimica et Cosmochimica Acta* 61: 4915–4931.

Borg LE, Shearer CK, Nyquist LE, and Norman M (2002) Isotopic constraints on the origin of lunar ferroan anorthosites. In: *Proceedings of the Lunar and Planetary Science Conference*, vol. 33, Abstract No. 1396.

Bougher SW, Hunten DM, and Phillips RJ (eds.) (1997) *Venus II – Geology, Geophysics, Atmosphere, and Solar Wind Environment*. In: *Space Science Series*p. 1330. Tucson, AZ: University of Arizona Press.

Braginsky SI (1964) Magnetohydrodynamics of the Earth core. *Geomagnetism and Aeronomy* 4: 698–712.

Brandon AD, Walker RJ, Morgan JW, and Goles GG (2000) Re–Os isotopic evidence for early differentiation of the Martian mantle. *Geochimica et Cosmochimica Acta* 64: 4083–4095.

Breuer D, Hauck S, Buske M, Pauer M, Spohn T, and Solomon S (2007) Interior evolution of Mercury. *Space Science Reviews* 132(2–4): 229–260. http://dx.doi.org/10.1007/s11214-007-9228-9.

Breuer D, Labrosse S, and Spohn T (2010) Thermal evolution and magnetic field generation in terrestrial planets and satellites. *Space Science Reviews* 152(1): 449–500.

Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research: Planets* 108(E7): 5072. http://dx.doi.org/10.1029/20002JE001999.

Breuer D and Spohn T (2006) Viscosity of the Martian mantle and its initial temperature: Constraints from crust formation history and the evolution of the magnetic field. *Planetary and Space Science* 54: 153–169.

Breuer D, Spohn T, and Wüllner U (1993) Mantle differentiation and the crustal dichotomy of Mars. *Planetary and Space Science* 41: 269–283.

Breuer D, Yuen DA, Spohn T, and Zhang S (1998) Three-dimensional models of Martian mantle convection with phase transitions. *Geophysical Research Letters* 25: 229–232.

Brown CD and Grimm RE (1996) Lithospheric rheology and flexure at Artemis Chasma, Venus. *Journal of Geophysical Research* 101(E5): 12697–12708.

Bruntt H, Deleuil M, Fridlund M, et al. (2010) Improved stellar parameters of CoRoT-7. arXiv, preprint arXiv:1005.3208.

Burov E and Diament M (1995) The effective elastic thickness (T_e) of the continental lithosphere: What does it really mean? *Journal of Geophysical Research* 100: 3905–3927.

Buske M (2006) *Dreidimensionale thermische Evolutionsmodelle für das Innere von Mars und Merkur*. PhD Thesis, Copernicus GmbH. ISBN 3-936586-58-6.

Buske M and Christensen U (2003) Three-dimensional evolution model of convection in the martian mantle. In: *EGS–AGU–EUG Joint Assembly*, Nice, France, 6–11 April 2003, Abstract #8497.

Busse FH, Richards AA, and Lenardic A (2006) A simple model of high Prandtl and high Rayleigh number convection bounded by thin low-viscosity layers. *Geophysical Journal International* 164(1): 160–167.

Byrne PK, Klimczak C, Şengör AMC, Solomon SC, Watters TR, and Hauck SA II (2014) Mercury's global contraction much greater than earlier estimates. *Nature Geoscience* 7: 301–307.

Cadek O and Fleitout L (2003) Effect of lateral viscosity variations in the top 300 km on the geoid and dynamic topography. *Geophysical Journal International* 152(3): 566–580. http://dx.doi.org/10.1046/j.1365-246X.2003.01859.x.

Cameron AGW, Fegley B Jr., Benz W, and Slattery WL (1988) The strange density of Mercury: Theoretical considerations. In: Vilas F, et al. (eds.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.

Campbell BA, Campbell DB, and DeVries CH (1999) Surface processes in the Venus highlands: Results from analysis of Magellan and Arecibo data. *Journal of Geophysical Research* 104(E1): 1897–1916.

Carter JA, Agol E, Chaplin WJ, et al. (2012) Kepler-36: A pair of planets with neighboring orbits and dissimilar densities. *Science* 337(6094): 556–559.

Cassen P and Reynolds RT (1973) Role of convection in the moon (solid state convection role in moon from analysis of models with homogeneous initial distribution of radioactive heat sources). *Journal of Geophysical Research* 78: 3203–3215.

Cassen P, Reynolds RT, Graziani F, Summers AL, McNellis J, and Blalock L (1979) Convection and lunar thermal history. *Physics of the Earth and Planetary Interiors* 19: 183.

Cassen P, Young RE, Schubert G, and Reynolds RT (1976) Implications of an internal dynamo for the thermal history of Mercury. *Icarus* 28: 501–508.
Charlier B, Grove TL, and Zuber MT (2013) Phase equilibria of ultramafic compositions on Mercury and the origin of the compositional dichotomy. *Earth and Planetary Science Letters* 363: 50–60.
Chen B, Li J, and Hauck SA II (2008) Non-ideal liquidus curve in the Fe–S system and Mercury's snowing core. *Geophysical Research Letters* 35: L07201. http://dx.doi.org/10.1029/2008GL033311.
Chen JH and Wasserburg GJ (1986) Formation ages and evolution of Shergotty and its parent planet from U–Th–Pb systematics. *Geochimica et Cosmochimica Acta* 50: 955–968.
Choblet G and Sotin C (2000) 3D thermal convection with variable viscosity: Can transient cooling be described by a quasi-static scaling law? *Physics of the Earth and Planetary Interiors* 119(3–4): 321–336.
Chopelas A and Boehler R (1989) Thermal expansion measurement at very high pressure, systematics, and a case for a chemically homogeneous mantle. *Geophysical Research Letters* 16: 1347.
Christensen UR (1984) Convection with pressure- and temperature-dependent non-Newtonian rheology. *Geophysical Journal of the Royal Astronomical Society* 77: 343–384. http://dx.doi.org/10.1111/j.1365-246X.1984.tb01939.x.
Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.
Christensen UR and Wicht J (2007) Numerical dynamo simulations. In: Schubert G (ed.) *Treatise on Geophysics, Vol. 8: Core Dynamics*, pp. 254–282. Amsterdam: Elsevier.
Christensen PR, Bandfield JL, Clark RN, Edgett KS, Hamilton VE, Hoefen T, et al. (2000) Detection of crystalline hematite mineralization on Mars by the Thermal Emission Spectrometer: Evidence for nearsurface water. *Journal of Geophysical Research* 105(E4): 9623–9642. http://dx.doi.org/10.1029/1999JE001093.
Chudinovskikh L and Boehler R (2007) Eutectic melting in the system Fe–S to 44 GPa. *Earth and Planetary Science Letters* 257(1): 97–103.
Cisowski SM and Fuller M (1986) Lunar paleointensities via the IRMs normalization method and the early magnetic history of the moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 411–424. Houston: Lunar and Planetary Science Institute.
Clauser C and Huenges E (1995) Thermal conductivity of rocks and minerals. *Rock Physics and Phase Relations: A Handbook of Physical Constants*. *AGU Reference Shelf*, vol. 3, Washington, DC: American Geophysical Union.
Clifford SM (1993) The hydrologic response of Mars to the onset of a colder climate and to the thermal evolution of its early crust. In: *Workshop on Early Mars: How Warm and How Wet?* Part 1, pp. 7–9.
Clifford SM and Fanale FP (1985) The thermal conductivity of the Martian crust. *Lunar and Planetary Sciences* XVI: 144–145.
Coleman PJ Jr and Russell CT (1977) The remanent magnetic field of the moon. *Royal Society of London Philosophical Transactions Series A* 285: 489–506.
Collinson DW (1997) Magnetic properties of Martian meteorites: Implications for an ancient Martian magnetic field. *Meteoritics* 803: 32–36.
Condie KC (1997) *Plate Tectonics and Crustal Evolution*, 4th edn. Oxford: Butterworth-Heinemann, 282 pp.
Connerney J, Acuña MH, and Ness NF (2003) Magnetic mapping of Mars: Year two. In: *EGS–AGU–EUG Joint Assembly*, Nice, France, 6–11 April 2003, Abstract #4445.
Connerney JEP, Acuña MH, Ness NF, et al. (2005) From the cover: Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102: 14970–14975.
Connerney JEP, Acuña MH, Wasilewski P, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
Connerney JEP and Ness NF (1988) Mercury's magnetic field and interior. In: *Mercury*, pp. 494–513. Tucson, AZ: University of Arizona Press.
Dauphas N and Pourmand A (2011) Hf–W–Th evidence for rapid growth of Mars and its status as a planetary embryo. *Nature* 473(7348): 489–492.
Davaille A and Jaupart C (1993) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
de Koker N (2010) Thermal conductivity of MgO periclase at high pressure: Implications for the D″ region. *Earth and Planetary Science Letters* 292: 392–398.
de Koker N, Steinle-Neumann G, and Vlček V (2012) Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T, and heat flux in Earth's core. *Proceedings of the National Academy of Sciences* 109(11): 4070–4073.
Debaille V, Brandon AD, O'Neill C, Yin QZ, and Jacobsen B (2009) Early martian mantle overturn inferred from isotopic composition of nakhlite meteorites. *Nature Geoscience* 2: 548–552.
Debaille V, Brandon AD, Yin Q-Z, and Jacobsen B (2007) Coupled 142Nd-143Nd evidence for a protracted magma ocean in Mars. *Nature* 450: 525–528.
Deschamps F and Sotin C (2000) Inversion of two-dimensional numerical convection experiments for a fluid with a strongly temperature-dependent viscosity. *Geophysical Journal International* 143: 204–218.
Deschamps F, Tackley PJ, and Nakagawa T (2010) Temperature and heat flux scalings for isoviscous thermal convection in spherical geometry. *Geophysical Journal International* 182: 137–154.
de Vries J, van den Berg A, and van Westrenen W (2010) Formation and evolution of a lunar core from ilmenite-rich magma ocean cumulates. *Earth and Planetary Science Letters* 292(1): 139–147.
Di Achille G, Popa C, Massironi M, et al. (2012) Mercury's radius change estimates revisited using MESSENGER data. *Icarus* 221: 456–460.
Dietrich W and Wicht J (2013) A hemispherical dynamo model: Implications for the Martian crustal magnetization. *Physics of the Earth and Planetary Interiors* 217: 10–21.
Dreibus G and Wänke H (1985) Mars: A volatile rich planet. *Meteoritics* 20: 367–382.
Dunlop DJ and Özdemir Ö (1997) *Rock Magnetism – Fundaments and Frontiers*. Cambridge: Cambridge University Press.
Dwyer CA, Stevenson DJ, and Nimmo F (2011) A long-lived lunar dynamo driven by continuous mechanical stirring. *Nature* 479: 212–214. http://dx.doi.org/10.1038/nature10564.
Dyal P, Parkin CW, and Daily WD (1974) Magnetism and the interior of the Moon. *Reviews of Geophysics and Space Physics* 12: 568–591.
Elkins-Tanton LT (2008) Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth and Planetary Science Letters* 271: 181–191.
Elkins-Tanton LT (2012) Magma oceans in the inner solar system. *Annual Review of Earth and Planetary Sciences* 40: 113–139.
Elkins-Tanton LT, Burgess S, and Yin Q-Z (2011) The lunar magma ocean: Reconciling the solidification process with lunar petrology and geochronology. *Earth and Planetary Science Letters* 304: 326–336.
Elkins-Tanton LT and Grove TL (2011) Water (hydrogen) in the lunar mantle: Results from petrology and magma ocean modeling. *Earth and Planetary Science Letters* 307: 173–179.
Elkins-Tanton LT, Parmentier EM, and Hess PC (2003) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38(12): 1753–1771.
Elkins-Tanton LT, Hess PC, and Parmentier EM (2005) Possible formation of ancient crust on Mars through magma ocean processes. *Journal of Geophysical Research* 110: E12S01, http://dx.doi.org/10.1029/2005JE002480.
Elkins-Tanton LT, Van Orman JA, Hager BH, and Grove TL (2002) Reexamination of the lunar magma ocean cumulate overturn hypothesis: Melting or mixing is required. *Earth and Planetary Science Letters* 196: 239–249.
Fegley B (2004) Venus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, pp. 487–507. Elsevier: Amsterdam.
Fegley B Jr. and Cameron AGW (1987) A vaporization model for iron/silicate fractionation in the Mercury protoplanet. *Earth and Planetary Science Letters* 82: 207–222.
Fegley B and Prinn RG (1989) Estimation of the rate of volcanism on Venus from reaction rate measurements. *Nature* 337: 55–58.
Fei Y, Bertka CM, and Finger LW (1997) High-pressure iron-sulfur compound, Fe_3S_2, and melting relations in the Fe–FeS system. *Science* 275(5306): 1621–1623.
Fei Y, Li J, Bertka CM, and Prewitt CT (2000) Structure type and bulk modulus of Fe_3S, a new iron-sulfur compound. *American Mineralogist* 85(11–12): 1830–1833.
Filiberto J and Treiman AH (2009) Martian magmas contained abundant chlorine, but little water. *Geology* 37(12): 1087–1090.
Fischer HJ and Spohn T (1990) Thermal-orbital histories of viscoelastic models of Io (J1). *Icarus* 83: 39–65.
Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.
Forsyth DW and Uyeda S (1975) On the relative importance of the driving forces of plate motions. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200.
Fowler AC and O'Brien SBG (1996) A mechanism for episodic subduction on Venus. *Journal of Geophysical Research* 101: 4755–4763.
Fraeman AA and Korenaga J (2010) The influence of mantle melting on the evolution of Mars. *Icarus* 210: 43–57.
Frey HV, Roark JH, Shockey KM, Frey EL, and Sakimoto SEH (2002) Ancient lowlands on Mars. *Geophysical Research Letters* (29). http://dx.doi.org/10.1029/2001GL013832.
Frey H and Schultz RA (1988) Large impact basins and the mega-impact origin for the crustal dichotomy on Mars. *Geophysical Research Letters* 15: 229–232.
Fuller M and Cisowski S (1987) Lunar paleomagnetism. In: Jacobs J (ed.) *Geomagnetism*, pp. 307–456. San Diego, CA: Academic Press.

Gagnepain-Beyneix J, Lognonné P, Chenet H, Lombardi D, and Spohn T (2006) A seismic model of the lunar mantle and constraints on temperature and mineralogy. *Physics of the Earth and Planetary Interiors* 159(3): 140–166.
Garcia R, Gagnepain-Beyneix J, Chevrot S, and Lognonné P (2011) Very preliminary reference Moon model. *Physics of the Earth and Planetary Interiors* 188: 96–113.
Garrick-Bethell I, Weiss BP, Shuster DL, and Buz J (2009) Early lunar magnetism. *Science* 323: 356–359. http://dx.doi.org/10.1126/science.1166804.
Ghias SR and Jarvis GT (2008) Mantle convection models with temperature- and depth-dependent thermal expansivity. *Journal of Geophysical Research* 113: B08408. http://dx.doi.org/10.1029/2007JB005355.
Ghosh A and McSween HY (1998) A thermal model for the differentiation of asteroid 4 Vesta, based on radiogenic heating. *Icarus* 134(2): 187–206.
Giampieri G and Balogh A (2002) Mercury's thermoelectric dynamo model revisited. *Planetary and Space Science* 50: 757–762.
Golabek GJ, Gerya TV, Kaus BJP, Ziethe R, and Tackley PJ (2009) Rheological controls on the terrestrial core formation mechanism. *Geochemistry, Geophysics, Geosystems* 10(11).
Golabek GJ, Keller T, Gerya TV, Zhu G, Tackley PJ, and Connolly JA (2011) Origin of the martian dichotomy and Tharsis from a giant impact causing massive magmatism. *Icarus* 215(1): 346–357.
Goldsmith JR (1982) Plagioclase stability at elevated temperatures and water pressures. *American Mineralogist* 67: 653–675.
Gomi H, Ohta K, Hirose K, et al. (2013) The high conductivity of iron and thermal evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 224: 88–103.
Goncharov A, Haugen B, Struzhkin V, Beck P, and Jacobsen S (2008) Radiative conductivity in the Earth's lower mantle. *Nature* 456(7219): 231–234.
Grasset O and Parmentier EM (1998) Thermal convection in a volumetrically heated, infinite Prandtl number fluid with strongly temperature-dependent viscosity: Implications for planetary thermal evolution. *Journal of Geophysical Research* 103: 18171–18181.
Greeley R and Schneid BD (1991) Magma generation on Mars: Amounts, rates, and comparisons with Earth, Moon and Venus. *Science* 254: 996–999.
Greeley R and Spudis PD (1981) Volcanism on Mars. *Reviews of Geophysics and Space Physics* 19: 13–41.
Grimm RE (1994) The deep structure of Venusian plateau highlands. *Icarus* 112(1): 89–103.
Grimm RE (2013) Geophysical constraints on the lunar Procellarum KREEP Terrane. *Journal of Geophysical Research* 118: 768–778. http://dx.doi.org/10.1029/2012JE004114.
Grossman L (1972) Condensation in the primitive solar nebula. *Geochimica et Cosmochimica Acta* 36: 597–619.
Grott M, Baratoux D, Hauber E, et al. (2013) Long-term evolution of the Martian crust-mantle system. *Space Science Review* 172(1): 49–111. http://dx.doi.org/10.1007/s11214-012-9948-3.
Grott M and Breuer D (2008) The evolution of the martian elastic lithosphere and implications for crustal and mantle rheology. *Icarus* 193: 503–515. http://dx.doi.org/10.1016/j.icarus.2007.08.015.
Grott M and Breuer D (2009) Implications of large elastic thicknesses for the composition and current thermal state of Mars. *Icarus* 201(2): 540–548.
Grott M and Breuer D (2010) On the spatial variability of the Martian elastic lithosphere thickness: Evidence for Mantle plumes? *Journal of Geophysical Research* 115: E03005. http://dx.doi.org/10.1029/2009JE003456.
Grott M, Breuer D, and Laneuville M (2011a) Thermo-chemical evolution and global contraction of Mercury. *Earth and Planetary Science Letters* 307: 135–146. http://dx.doi.org/10.1016/j.epsl.2011.04.040.
Grott M, Hauber V, Werner SC, Kronberg P, and Neukum G (2005) High heat flux on ancient Mars: Evidence from rift flank uplift at Coracis Fossae. *Geophysical Research Letters* 32: L21201. http://dx.doi.org/10.1029/2005GL023894.
Grott M, Hauber E, Werner SC, Kronberg P, and Neukum G (2007) Mechanical modeling of thrust faults in the Thaumasia region, Mars, and implications for the Noachian heat flux. *Icarus* 186: 517–526. http://dx.doi.org/10.1016/j.icarus.2006.10.001.
Grott M, Morschhauser A, Breuer D, and Hauber E (2011b) Volcanic outgassing of CO_2 and H_2O on Mars. *Earth and Planetary Science Letters* 308(3–4): 391–400. http://dx.doi.org/10.1016/j.epsl.2011.06.014.
Gurnis M (1986) The effects of chemical density differences on convective mixing in the Earth's mantle. *Journal of Geophysical Research* 91: 11407.
Hager BH and Richards MA (1989) Long-wavelength variations in earth's geoid – Physical models and dynamical implications. *Philosophical Transactions of the Royal Society A: Physical, Mathematical and Engineering Sciences* 328: 309–327.
Halliday AN, Wänke H, Birck J-L, and Clayton RN (2001) The accretion, composition and early differentiation of Mars. *Space Science Reviews* 96: 197–230.
Hamilton CW, Beggan CD, Still S, et al. (2013) Spatial distribution of volcanoes on Io: Implications for tidal heating and magma ascent. *Earth and Planetary Science Letters* 361: 272–286.
Hansen U, Yuen DA, Kroening SE, and Larsen TB (1993) Dynamical consequences of depth-dependent thermal expansivity and viscosity on mantle circulations and thermal structure. *Physics of the Earth and Planetary Interiors* 77: 205–223.
Harder H (1998) Phase transitions and the 3D-planform of thermal convection in the Martian mantle. *Journal of Geophysical Research* 103: 16775–16798.
Harder H (2000) Mantle convection and the dynamic geoid of Mars. *Geophysical Research Letters* 27: 301–304.
Harder H and Christensen U (1996) A one-plume model of Martian mantle convection. *Nature* 380: 507–509.
Harder H and Schubert G (2001) Sulfur in Mercury's core? *Icarus* 151: 118–122.
Harper CL Jr, Nyquist LE, Bansal B, Wiesmann H, and Shih CY (1995) Rapid accretion and early differentiation of Mars indicated by 142Nd/144Nd in SNC meteorites. *Science* 267(5195): 213–217.
Hartmann WK, Malin M, McEwen A, et al. (1999) Evidence for recent volcanism on Mars from crater counts. *Nature* 397: 586–589.
Hartmann WK and Neukum W (2001) Cratering chronology and the evolution of Mars. *Space Science Reviews* 96(1/4): 165–194.
Hatzes AP, Fridlund M, Nachmani G, et al. (2011) The mass of corot-7b. *Astrophysical Journal* 743(1): 75.
Hauck SA, Aurnou JM, and Dombard AJ (2006) Sulfur's impact on core evolution and magnetic field generation on ganymede. *Journal of Geophysical Research* 111.
Hauck SA, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222(3–4): 713–728.
Hauck SA II, Margot J-L, Solomon SC, et al. (2013) The curious case of Mercury's internal structure. *Journal of Geophysical Research* 118: 1303–1322.
Hauck SA and Phillips RJ (2002) Thermal and crustal evolution of Mars. *Journal of Geophysical Research* 107(E7). http://dx.doi.org/10.1029/2001JE001801.
Head JW, Chapman CR, Strom RG, Fassett CI, Denevi BW, Blewett DT, et al. (2011) Flood Volcanism in the Northern High Latitudes of Mercury Revealed by MESSENGER. *Science* 333: 1853.
Head JW, Greeley R, Golombek MP, et al. (2001) Geological processes and evolution. *Space Science Reviews* 96(1/4): 263–292.
Head J and Hiesinger H (1999) Lunar mare stratigraphy and lunar evolution. *Geophysical Research Abstracts* 1: 735.
Heimpel MH, Aurnou JM, Al-Shamali FM, and Gomez Perez N (2005) A numerical study of dynamo action as function of spherical shell geometry. *Earth and Planetary Science Letters* 236: 542–557.
Helbert J and Benkhoff J (2003) A new approach to assessing the burial depth of ground ice deposits and its application to proposed MER landing sites in Isidis Planitia. *Journal of Geophysical Research* 108(E12), ROV 28–31.
Helbert J, Mariangeli L, Piccioni G, Drossart P, and Mueller N (2006) The VIRTIS Venus Express Team, observing the surface of Venus with VIRTIS on Venus Express. American Astronomical Society, DPS Meeting #38, #16.03, 2006.
Herrick RR (1994) Resurfacing history of Venus. *Geology* 22: 703–706.
Herrick RR and Parmentier EM (1994) Episodic large-scale overturn of two-layer mantle in terrestrial planets. *Journal of Geophysical Research* 99: 2053–2062.
Hess PC (2000) Petrogenesis of lunar troctolites – Implications for the Moon and its evolution. In: *31st Annual Lunar and Planetary Science Conference*, Houston, TX, 13–17 March 2000, Abstract No. 1389.
Hess PC and Parmentier EM (1995) A model for the thermal and chemical evolution of the Moon's interior: Implications for the onset of mare volcanism. *Earth and Planetary Science Letters* 134(3–4): 501–514.
Hess PC and Parmentier EM (2001) Implications of magma ocean cumulate overturn for Mars. In: *32nd Annual Lunar and Planetary Science Conference*, Houston, TX, 12–16 March 2001, Abstract No. 1319.
Hide R (1972) Comments on the Moon's magnetism. *Moon* 4: 39.
Hiesinger H, Head JW III, Wolf U, Jaumann R, and Neukum G (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108: 5065. http://dx.doi.org/10.1029/2002JE001985.
Hiesinger H, Jaumann R, Neukum G, and Head JW (2000) Ages of mare basalts on the lunar nearside. *Journal of Geophysical Research* 105: 29239–29275.
Hirth G and Kohlstedt DL (1996) Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144(1): 93–108.
Hofmeister AM (1999) Mantle values of thermal conductivity and the geotherm from phonon lifetimes. *Science* 283(5408): 1699.

Hofmeister AM (2008) Inference of high thermal transport in the lower mantle from laser-flash experiments and the damped harmonic oscillator model. *Physics of the Earth and Planetary Interiors* 170: 201–206.

Hoogenboom T and Smrekar SE (2006) Elastic thickness estimates for the northern lowlands of Mars. *Earth and Planetary Science Letters* 248(3–4): 830–839.

Höink T, Lenardic A, and Richards M (2012) Depth-dependent viscosity and mantle stress amplification: Implications for the role of the asthenosphere in maintaining plate tectonics. *Geophysical Journal International* 191: 30–41.

Hood L and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins: Two dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.

Hood LL, Herbert F, and Sonett CP (1982) Further efforts to limit lunar internal temperatures from electrical conductivity determinations. *Journal of Geophysical Research* 87(S01): A109–A116. http://dx.doi.org/10.1029/JB087iS01p0A109.

Hood L and Vickery A (1984) Magnetic field amplification and generation in hypervelocity meteoroid impacts with application to lunar paleomagnetism. In: *Proceedings of the 15th Lunar and Planetary Science Conference, Part 1. Journal of Geophysical Research* 89(supplement): C211–C223.

Hood LL, Zakharian A, Halekas J, et al. (2001) Initial mapping and interpretation of lunar crustal magnetic anomalies using Lunar Prospector magnetometer data. *Journal of Geophysical Research* 106: 27825–27839.

Howard AW, Sanchis-Ojeda R, Marcy GW, et al. (2013) A rocky composition for an Earth-sized exoplanet. *Nature* 503(7476): 381–384.

Humayun M, Nemchin A, Zanda B, et al. (2013) Origin and age of the earliest Martian crust from meteorite NWA 7533. *Nature* 503: 513–516.

Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171(2): 391–410.

Hüttig C and Breuer D (2011) Regime classification and planform scaling for internally heated mantle convection. *Physics of the Earth and Planetary Interiors* 186(3–4): 111–124. http://dx.doi.org/10.1016/j.pepi.2011.03.011.

Ishihara Y, Goossens S, Matsumoto K, Noda H, Araki H, Namiki N, et al. (2009) Crustal thickness of the Moon: Implications for farside basin structures. *Geophysical Research Letters* 36: L19202. http://dx.doi.org/10.1029/2009GL039708.

Ito K and Kennedy GC (1971) An experimental study of the basalt-garnet granulite-eclogite transition. In: Heacock JG (ed.) *The Structure and Physical Properties of the Earth's Crust*. *Geophysical Monograph Series*, vol. 14, pp. 303–314. Washington, DC: American Geophysical Union.

Ivanov MA and Head JW (2006) Testing directional (evolutionary) and non-directional models of the geologic history of Venus: Results of mapping in a geotraverse along the equator of Venus. In: *37th Annual Lunar and Planetary Science Conference*, League City, TX, 13–17 March 2006, Abstract No. 1366.

Jaeger WL, Turtle EP, Keszthelyi LP, Radebaugh J, and McEwen AS (2003) Orogenic tectonism on Io. *Journal of Geophysical Research* 108(E8), Art. no. 5093.

Jagoutz E (1991) Chronology of SNC meteorites. *Space Science Reviews* 56: 13–22.

Jarvis GT (1984) Time-dependent convection in the Earth's mantle. *Physics of the Earth and Planetary Interiors* 36: 305–327.

Johnson CL and Richards MA (2003) A conceptual model for the relationship between coronae and large-scale mantle dynamics on Venus. *Journal of Geophysical Research: Planets* 108(E6).

Johnson CL and Sandwell DT (1994) Lithospheric flexure on Venus. *Geophysical Journal International* 119: 627–647.

Jurdy DM and Stefanick M (2004) Vertical extrapolation of Mars magnetic potentials. *Journal of Geophysical Research* 109(E10), CiteID E10005.

Karato S (2013) Rheological properties of minerals and rocks. In: Karato S (ed.) *Physics and Chemistry of the Deep Earth*, pp. 94–144. Oxford: Wiley-Blackwell.

Karato S, Paterson MS, and Fitzgerald JD (1986) Rheology of synthetic olivine aggregates – Influence of grain size and water. *Journal of Geophysical Research* 91: 8151–8176.

Karato S and Rubie DC (1997) Towards an experimental study of deep mantle rheology: A new multi-anvil specimen assembly for deformation studies under high pressure and temperatures. *Journal of Geophysical Research* 102: 20111–20122.

Karato S and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.

Katsura T, Shatskiy A, Manthilake M, et al. (2009) Thermal expansion of forsterite at high pressures determined by in situ X-ray diffraction: the adiabatic geotherm in the upper mantle. *Physics of the Earth and Planetary Interiors* 174(1): 86–92.

Katsura T, Yokoshi S, Kawabe K, et al. (2009) P–V–T relations of $MgSiO_3$ perovskite determined by in situ X-ray diffraction using a large-volume high-pressure apparatus. *Geophysical Research Letters* 36: L01305. http:// dx.doi.org/10.1029/2008GL035658.

Kaufmann G and Lambeck K (2000) Mantle dynamics, postglacial rebound and the radial viscosity profile. *Physics of the Earth and Planetary Interiors* 121: 301–324.

Kaula WM, Schubert G, Lingenfelter RE, Sjogren WL, and Wollenhaupt WR (1972) Analysis and interpretation of lunar laser altimetry. In: *Proceedings of the Lunar Science Conference*, vol. 2, p. 2189.

Ke Y and Solomatov VS (2006) Early transient superplumes and the origin of the Martian crustal dichotomy. *Journal of Geophysical Research* (111). http://dx.doi.org/10.1029/2005JE002631.

Ke Y and Solomatov VS (2009) Coupled core–mantle thermal evolution of early Mars. *Journal of Geophysical Research: Planets* 114(13): 1–12.

Keller T and Tackley PJ (2009) Towards self-consistent modeling of the martian dichotomy: The influence of one-ridge convection on crustal thickness distribution. *Icarus* 202: 429.

Kemp AIS and Hawkesworth CJ (2003) Granitic perspectives on the generation and secular evolution of the continental crust. In: Rudnick RL (ed.) *The Crust, Vol. 3: Treatise on Geochemistry* (Holland HD and Turekian KK, eds.), pp. 349–410. Oxford: Elsevier-Pergamon. ISBN 0-08-043751-6.

Kerber L, Head JW, Blewett DT, et al. (2011) The global distribution of pyroclastic deposits on Mercury: The view from MESSENGER flybys 1–3. *Planetary and Space Science* 59(15): 1895–1909.

Khan A, Connolly JAD, Maclennan J, and Mosegaard K (2007) Joint inversion of seismic and gravity data for lunar composition and thermal state. *Geophysical Journal International* 168(1): 243–258.

Khurana KK, Jia X, Kivelson MG, Nimmo F, Schubert G, and Russell CT (2011) Evidence of a global magma ocean in Io's interior. *Science* 332(6034): 1186–1189.

Kiefer WS, Richards MA, Hager BH, and Bills BG (1986) A dynamic model of Venus's gravity field. *Geophysical Research Letters* 13: 14–17.

Kieffer HH (1976) Soil and surface temperatures at the Viking landing sites. *Science* 194: 1344–1346.

King SD (2008) Pattern of lobate scarps on Mercury's surface reproduced by a model of mantle convection. *Nature Geoscience* 1: 229–232. http://dx.doi.org/10.1038/ngeo152.

Kirk RL and Stevenson DJ (1989) The competition between thermal contraction and differentiation in the stress history of the Moon. *Journal of Geophysical Research* 94(B9): 12133–12144.

Kivelson MG, Khurana KK, Russell CT, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.

Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.

Klima R, Cahill J, Hagerty J, and Lawrence D (2013) Remote detection of magmatic water in Bullialdus Crater on the Moon. *Nature Geoscience* 6: 737–741. http://dx.doi.org/10.1038/ngeo1909.

Kobayashi S, Hasebe N, Shibamura E, et al. (2010) Determining the absolute abundances of natural radioactive elements on the lunar surface by the Kaguya gamma-ray spectrometer. *Space Science Reviews* 154(1–4): 193–218.

Konopliv AS and Sjogren WL (1994) Venus spherical harmonic gravity model to degree and order 60. *Icarus* 112(1): 42–54.

Konrad W and Spohn T (1997) Thermal history of the Moon: Implications for an early core dynamo and post-accretional magmatism. *Advances in Space Research* 19(10): 1511–1521.

Kronberg P, Hauber E, Grott M, Werner SC, Schäfer T, and Gwinner K (2007) Acheron Fossae, Mars: Tectonic rifting, volcanism, and implications for lithospheric thickness. *Journal of Geophysical Research* 112: E04005. http://dx.doi.org/10.1029/2006JE002780.

Kucinskas AB and Turcotte DL (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112(1): 104–116.

Kuwayama Y and Hirose K (2004) Phase relations in the system Fe-FeSi at 21 GPa. *American Mineralogist* 89: 273–276.

Lainey V, Arlot JE, Karatekin O, and van Hoolst T (2009) Strong tidal dissipation in Io and Jupiter from astrometric observations. *Nature* 459: 957–959.

Laneuville M, Wieczorek M, Breuer D, and Tosi N (2013) Asymmetric thermal evolution of the Moon. *Journal of Geophysical Research* 118: 1435–1452. http://dx.doi.org/10.1002/jgre.20103.

Langlais BM, Purucker E, and Mandea M (2004) Crustal magnetic field of Mars. *Journal of Geophysical Research* 109: E02008. http://dx.doi.org/10.1029/2003JE002048.

Lawrence DJ, Feldman WC, Barraclough BL, et al. (1998) Global element maps of the Moon: The Lunar Prospector gamma-ray spectrometer. *Science* 281: 1484–1489.

Lawrence DJ, Elphic RC, Feldman WC, Prettyman TH, Gasnault O, and Maurice S (2003) Small-area thorium features on the lunar surface. *Journal of Geophysical Research* 108 (E), 5102.

Lawrence K, Johnson C, Tauxe L, and Gee J (2008) Lunar paleointensity measurements: Implications for lunar magnetic evolution. *Physics of the Earth and Planetary Interiors* 168(1–2): 71–87.

Le Bars M, Wieczorek MA, Karatekin O, Cebron D, and Laneuville M (2011) An impact-driven dynamo for the early Moon. *Nature* 479: 215–218. http://dx.doi.org/10.1038/nature10565.

Lee D-C, Halliday AN, Snyder GA, and Taylor LA (1997) Age and origin of the Moon. *Science* 278(5340): 1098.

Lee T, Papanastassiou D, and Wasserburg G (1976) Demonstration of ^{26}Mg excess in Allende and evidence for ^{26}Al. *Geophysical Research Letters* 3: 109–112.

Léger A, Rouan D, Schneider J, et al. (2009) Transiting exoplanets from the CoRoT space mission VIII. CoRoT-7b: The first super-Earth with measured radius. arXiv, preprint arXiv:0908.0241.

Leitch AM, Yuen DA, and Sewell G (1991) Mantle convection with internal heating and pressure-dependent thermal expansivity. *Earth and Planetary Science Letters* 102: 213–232.

Lenardic A, Jellinek AM, and Moresi L-N (2008) A climate induced transition in the tectonic style of a terrestrial planet. *Earth and Planetary Science Letters* 271: 34–42.

Lenardic A, Richards MA, and Busse FH (2006) Depth-dependent rheology and the horizontal length scale of mantle convection. *Journal of Geophysical Research* 111(B7).

Leshin LA, Epstein S, and Stolper EM (1996) Hydrogen isotope geochemistry of SNC meteorites. *Geochimica et Cosmochimica Acta* 60(14): 2635–2650.

Lewis JS (1972) The temperature gradient in the solar nebula. *Science* 186: 440–443.

Li Q and Kiefer WS (2007) Mantle convection and magma production on present-day Mars: Effects of temperature-dependent rheology. *Geophysical Research Letters* 34: L16203. http://dx.doi.org/10.1029/2007GL030544.

Liang Y, Price JD, Wark DA, and Watson EB (2001) Nonlinear pressure diffusion in a porous medium: Approximate solutions with applications to permeability measurements using transient pulse decay method. *Journal of Geophysical Research* 106: 529–535.

Lillis RJ, Frey HV, and Manga M (2008) Rapid decrease in Martian crustal magnetization in the Noachian era: Implications for the dynamo and climate of early Mars. *Geophysical Research Letters* 35(14). http://dx.doi.org/10.1029/2008GL034338.

Lillis RJ, Robbins S, Manga M, Halekas JS, and Frey HV (2013) Time history of the Martian dynamo from crater magnetic field analysis. *Journal of Geophysical Research: Planets* 118(7): 1488–1511. http://dx.doi.org/10.1002/jgre.20105.

Lin RP, Mitchell DL, Curtis DW, et al. (1998) Lunar surface magnetic fields and their interaction with the solar wind: Results from Lunar Prospector. *Science* 281: 1480–1484.

Lister JR and Buffett BA (1995) The strength and efficiency of thermal and compositional convection in the geodynamo. *Physics of the Earth and Planetary Interiors* 91: 17–30.

Lodders K and Fegley B (1998) The Planetary Scientist's Companion. New York: Oxford University Press. QB601. L84

Loper DE and Werner C (2000) On the cause of lunar crustal asymmetries. In: *31st Annual Lunar and Planetary Science Conference*, Houston, TX, 13–17 March 2000, Abstract No. 1746.

Lucey PG, Blewett DT, and Jolliff BL (2000) Lunar iron and titanium abundance algorithms based on final processing of clementine ultraviolet-visible images. *Journal of Geophysical Research* 105(E8): 20297–20305. http://dx.doi.org/10.1029/1999JE001117.

Mackwell SJ (1991) High-temperature rheology of enstatite: Implications for creep in the mantle. *Geophysical Research Letters* 18: 2027–2030.

MacPherson GJ, Davis DM, and Zinner EK (1995) The distribution of aluminum-26 in the early solar system-A reappraisal. *Meteoritics* 30: 365–386.

Malavergne V, Toplis MJ, Berthet S, and Jones J (2010) Highly reducing conditions during core formation on Mercury: Implications for internal structure and the origin of a magnetic field. *Icarus* 206(1): 199–209.

Manglik A, Wicht J, and Christensen UR (2010) A dynamo model with double diffusive convection for Mercury's core. *Earth and Planetary Science Letters* 289(3–4): 619–628. http://dx.doi.org/10.1016/j.epsl.2009.12.007.

Margot JL, Peale SJ, Jurgens RF, Slade MA, and Holin IV (2007) Large longitude libration of Mercury reveals a molten core. *Science* 316(5825): 710–714. http://dx.doi.org/10.1126/science.1140514.

Marinova MM, Aharonson O, and Asphaug E (2008) Mega-impact formation of the Mars hemispheric dichotomy. *Nature* 453 (7199): 1216–1219. http://dx.doi.org/10.1038%2Fnature07070.

Matsui T and Abe Y (1986) Origin of the Moon and its early thermal evolution. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 453–468. Houston, TX: Lunar and Planetary Institute.

Matyska C and Yuen DA (2005) The importance of radiative heat transfer on superplumes in the lower mantle with the new post-perovskite phase change. *Earth and Planetary Science Letters* 234: 71–81.

McCallum IS (2001) A new view of the Moon in light of data from Clementine and Prospector missions. *Earth, Moon and Planets* 85–86: 253–269.

McCammon CA, Ringwood AE, and Jackson I (1983) Thermodynamic of the system Fe-FeO-MgO at high pressure and temperature and a model for the formation of the Earth's core. *Geophysical Journal of the Royal Astronomical Society* 72: 577–595.

McCubbin F, Francis M, Riner M, Vander Kaaden K, and Burkemper L (2012) Is Mercury a volatile-rich planet? *Geophysical Research Letters* 39(9): L09202.

McCubbin FM, Hauri EH, Elardo SM, Vander Kaaden KE, Wang J, and Shearer CK (2012) Hydrous melting of the martian mantle produced both depleted and enriched shergottites. *Geology* 40: 683–686.

McCubbin FM, Jolliff BL, Nekvasil H, Carpenter PK, Zeigler RA, Steele A, et al. (2011) Fluorine and chlorine abundances in lunar apatite: Implications for heterogeneous distributions of magmatic volatiles in the lunar interior. *Geochimica et Cosmochimica Acta* 75: 5073–5093.

McCubbin FM, Smirnov A, Nekvasil H, Wang J, Hauri E, and Lindsley DH (2010) Hydrous magmatism on Mars: A source of water for the surface and subsurface during the Amazonian. *Earth and Planetary Science Letters* 292: 132–138.

McEwen AS, Keszthelyi LP, Lopes R, Schenk PM, and Spencer JR (2004) The lithosphere and surface of Io. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 307–328. Cambridge: Cambridge University Press.

McEwen AS, Keszthelyi L, Spencer JR, et al. (1998) High-temperature silicate volcanism on Jupiter's moon Io. *Science* 281(5373): 87–90.

McGill GE and Dimitriou AM (1990) Origin of the Martian global dichotomy by crustal thinning in the late Noachian or early Hesperian. *Journal of Geophysical Research* 95: 12595–12605.

McGovern PJ, Solomon SC, Smith DE, et al. (2004) Correction to "Localized gravity/topography admittance and correction spectra on Mars: Implications for regional and global evolution" *Journal of Geophysical Research* 109(E7), CiteID E07007.

McKenzie D (1977) The initiation of trenches: A finite amplitude instability. In: Talwani M and Piman WC (eds.) *Island Arcs, Deep Sea Trenches and Back-Arc Basins. Maurice Ewing Series*, vol. 1, pp. 57–61. Washington, DC: American Geophysical Union.

McKenzie D (1994) The relationship between topography and gravity on Earth and Venus. *Icarus* 112(1): 55–88.

McKenzie D, Ford PG, Johnson C, et al. (1992) Features on Venus generated by plate boundary processes. *Journal of Geophysical Research* 97: 13533–13544.

McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130(1): 198–216.

McKinnon WB, Zahnle K, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Modeling and observations. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 969–1014. Tucson, AZ: University of Arizona Press.

McNutt MK (1984) Lithospheric flexure and thermal anomalies. *Journal of Geophysical Research* 89: 11180–11194.

McSween HY Jr. (1985) SNC meteorites: Clues to Martian petrologic evolution? *Reviews of Geophysics* 23: 391–416.

McSween HY, Grove TL, Lentz RCF, et al. (2001) Geochemical evidence for magmatic water within Mars from pyroxenes in Shergotty meteorite. *Nature* 409: 487–490.

McSween HY, Grove TL, and Wyatt MB (2003) Constraints on the composition and petrogenesis of the Martian crust. *Journal of Geophysical Research* 108(E12): 5125. http://dx.doi.org/10.1029/2003JE002175.

Mei S and Kohlstedt DL (2000) Influence of water on plastic deformation of olivine aggregates: 1. Diffusion creep regime. *Journal of Geophysical Research: Solid Earth (1978–2012)* 105(B9): 21457–21469.

Merk R, Breuer D, and Spohn T (2002) Numerical modeling of ^{26}Al-induced radioactive melting of asteroids considering accretion. *Icarus* 159: 183–191.

Meyer J, Elkins Tanton LT, and Wisdom J (2010) Coupled thermal-orbital evolution of the early Moon. *Icarus* 208: 1–10.

Michel NC, Hauck SA, Solomon SC, Phillips RJ, Roberts JH, and Zuber MT (2013) Thermal evolution of Mercury as constrained by MESSENGER observations. *Journal of Geophysical Research: Planets* 118: 1033–1044.

Michel N and Forni O (2011) Mars mantle convection: Influence of phase transitions with core cooling. *Planetary and Space Science* 59: 741–748.

Minear JW and Fletcher CR (1978) Crystallization of a lunar magma ocean. In: *Lunar and Planetary Science Conference, 9th*, Houston, TX, 13–17 March, vol. 1, pp. 263–283. New York: Pergamon Press.

Mitchell DL, Lillis RJ, Lin RP, Connerney JEP, and Acuña MH (2007) A global map of Mars' crustal magnetic field based on electron reflectometry. *Journal of Geophysical Research* 112: E01002. http://dx.doi.org/10.1029/2005JE002564.

Mitchell DL, Lin RP, Mazelle C, et al. (2001) Probing Mars' crustal magnetic field and ionosphere with the MGS electron reflectometer. *Journal of Geophysical Research* 106(E10): 23419–23428.

Mitchell DL, Lin RP, Rème H, Acuña MH, Cloutier PA, and Ness NF (1999) Crystal magnetospheres observed in the Martian night hemisphere. American Astronomical Society, DPS Meeting #31, Abstracts #59.04, *Bulletin of the American Astronomical Society*, 31, p. 1584.

Mohit PS and Arkani-Hamed J (2004) Impact demagnetization of the Martian crust. *Icarus* 168: 305–317.

Monteux J, Jellinek AM, and Johnson CL (2013) Dynamics of core merging after a mega-impact with applications to Mars' early dynamo. *Icarus* 226(1): 20–32.

Moore WB (2001) The thermal state of Io. *Icarus* 154: 548–550.

Moore WB (2003) Tidal heating and convection in Io. *Journal of Geophysical Research* 108: 5096. http://dx.doi.org/10.1029/2002JE001943.

Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regression of highland geoid and topography. *Icarus* 128(2): 415–428.

Moore WB and Webb AG (2013) Heat-pipe Earth. *Nature* 501: 501–505.

Morard G and Katsura T (2010) Pressure temperature cartography of the Fe-S-Si immiscible system. *Geochimica et Cosmochimica Acta* 74: 3659–3667.

Moresi L-N and Solomatov VS (1995) Numerical investigation of 2D convection with extremely large viscosity variations. *Physics of Fluids* 7: 2154–2162.

Moresi L-N and Solomatov VS (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic style of the Earth and Venus. *Geophysical Journal* 133: 669–682.

Morris S and Canright D (1984) A boundary-layer analysis of Benard convection in a fluid of strongly temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 36: 355–373.

Morschhauser A, Grott M, and Breuer D (2011) Crustal recycling, mantle dehydration, and the thermal evolution of Mars. *Icarus* 212(2): 541–558. http://dx.doi.org/10.1016/j.icarus.2010.12.028.

Mouginis-Mark PJ, Wilson L, and Zuber MT (1992) The physical volcanology of Mars. In: Jakosky BM, Snyder CW, Matthews MS, and Kieffer HH (eds.) *Mars*, pp. 424–452. Tucson, AZ: University of Arizona Press.

Murakami M, Hirose K, Sata N, Ohishi Y, and Kawamura K (2004) Phase transition of $MgSiO_3$ perovskite in the deep lower mantle. *Science* 304: 855–858.

Murthy VR, van Westrenen W, and Fei Y (2003) Experimental evidence that potassium is a substantial radioactive heat source in planetary cores. *Nature* 423: 163–165.

Nakagawa T and Tackley PJ (2012) Influence of magmatism on mantle cooling, surface heat flow and Urey ratio. *Earth and Planetary Science Letters* 329: 1–10.

Ness NF, Acuña MH, Connerney J, et al. (1999) MGS magnetic fields and electron reflectometer investigation: Discovery of paleomagnetic fields due to crustal remanence. *Advances in Space Research* 23(11): 1879–1886.

Ness NF, Behannon KW, Lepping RP, Schatten KH, and Whang YC (1974) Magnetic field observations near Mercury: Preliminary results from Mariner 10. *Science* 185: 151–160.

Ness NF, Behannon KW, Lepping RP, and Whang YC (1975) Magnetic field of Mercury. *Journal of Geophysical Research* 80(Part I): 2708–2716.

Ness NF, Behannon KW, Lepping RP, and Whang YC (1976) Observations of Mercury's magnetic field. *Icarus* 28: 479–488.

Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on Mars revealed by the High Resolution Stereo Camera. *Nature* 432: 971–979. http://dx.doi.org/10.1038/nature 03231.

Neumann W, Breuer D, and Spohn T (2012) Differentiation and core formation in accreting planetesimals. *Astronomy and Astrophysics* 543: 141.

Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E08002. http://dx.doi.org/10.1029/2004JE002262.

Nimmo F (2000) Dike intrusion as a possible cause of linear Martian magnetic anomalies. *Geology* 28: 391–394.

Nimmo F (2002) Crustal analysis of Venus from Magellan satellite observations at Atalanta Planitia, Beta Regio, and Thetis Regio. *Geology* 30: 987–990.

Nimmo F, Hart DS, Korycansky DG, and Agnor CB (2008) Implications of an impact origin for the Martian hemispheric dichotomy. *Nature* 453(7199): 1220–1223. http://dx.doi.org/10.1038/nature07025.

Nimmo F and McKenzie D (1996) Modelling plume-related uplift, gravity and melting on Venus. *Earth and Planetary Science Letters* 145(1–4): 109–123.

Nimmo F and McKenzie D (1998) Volcanism and tectonism on Venus. *Annual Review of Earth and Planetary Sciences* 26: 23–53.

Nimmo F and Stevenson D (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11979.

Nimmo F and Tanaka K (2005) Early crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161.

Nimmo F and Watters TR (2004) Depth of faulting on Mercury: Implications for heat flux and crustal and effective elastic thickness. *Geophysical Research Letters* 31: L02701. doi:02710.01029/02003GL018847.

Nittler LR, Starr RD, Weider SZ, et al. (2011) The major-element composition of Mercury's surface from MESSENGER X-ray spectrometry. *Science* 333: 1847–1950. http://dx.doi.org/10.1126/science.1211567.

Noack L, Breuer D, and Spohn T (2012) Coupling the atmosphere with the interior dynamics: Implications for the resurfacing of Venus. *Icarus* 217(2): 484–498. http://dx.doi.org/10.1016/j.icarus.2011.08.026.

Norman MD (2002) Thickness and composition of the Martian crust revisited: Implications of an ultradepleted mantle with a Nd isotopic composition like that of QUE94201. In: *33rd Annual Lunar and Planetary Science Conference*, Houston, TX, 11–15 March 2002, Abstract No. 1175.

Nozette S, Rustan P, Pleasance LP, et al. (1994) The Clementine mission to the Moon: Scientific overview. *Science* 266: 1835–1839.

O'Reilly TC and Davies GF (1981) Magma transport of heat on Io: A mechanism allowing a thick lithosphere. *Geophysical Research Letters* 8: 313–316.

Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ in Earth's D″ layer. *Nature* 430: 445–448.

Ogawa M and Yanagisawa T (2011) Numerical models of Martian mantle evolution induced by magmatism and solid-state convection beneath stagnant lithosphere. *Journal of Geophysical Research: Planets (1991–2012)* 116(E8).

Ogawa M and Yanagisawa T (2012) Two-dimensional numerical studies on the effects of water on Martian mantle evolution induced by magmatism and solid-state mantle convection. *Journal of Geophysical Research: Planets* 117(E6).

Ohta K, Yagi T, Taketoshi N, et al. (2012) Lattice thermal conductivity of $MgSiO_3$ perovskite and post-perovskite at the core-mantle boundary. *Earth and Planetary Science Letters* 349–350: 109–115.

Ohtake M, Takeda H, Matsunaga T, et al. (2012) Asymmetric crustal growth on the Moon indicated by primitive farside highland materials. *Nature Geoscience* 5: 384–388.

Palme H, Spettel B, Winke H, Bischoff A, and Stoeffler D (1984) Early differentiation of the Moon: Evidence from trace elements in plagioclase. In: *Proceedings of the 15th Lunar and Planetary Science Conference*, Houston, TX, 12–16 March 1984 (*Journal of Geophysical Research* 89: C3–Cl5).

Parmentier EM and Hess PC (1992) Chemical differentiation of a convecting planetary interior: Consequences for a one plate planet such as Venus. *Geophysical Research Letters* 19: 2015–2018.

Parmentier EM, Zhong S, and Zuber M (2002) Gravitational differentiation due to initial chemical stratification: Origin of lunar asymmetry by the creep of dense KREEP? *Earth and Planetary Science Letters* 201: 473–480.

Parmentier EM and Zuber MT (2007) Early evolution of Mars with mantle compositional stratification or hydrothermal crustal cooling. *Journal of Geophysical Research* 112: E02007. http://dx.doi.org/10.1029/2005JE002626.

Pauer M, Fleming K, and Cadek O (2006) Modeling the dynamic component of the geoid and topography of Venus. *Journal of Geophysical Research* 111: E11012. http://dx.doi.org/10.1029/2005JE002511.

Peale SJ, Cassen P, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.

Peltier WR and Jiang X (1996) Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *Journal of Geophysical Research* 101(B2): 3269–3290.

Pepe F, Cameron AC, Latham DW, et al. (2013) An Earth-sized planet with an Earth-like density. *Nature* 503(7476): 377–380.

Peplowski PN, Evans LG, Hauck SA, et al. (2011) Radioactive elements on Mercury's surface from MESSENGER: Implications for the planet's formation and evolution. *Science* 333: 1850–1852. http://dx.doi.org/10.1126/science.1211576.

Peplowski PN, Lawrence DJ, Rhodes EA, et al. (2012) Variations in the abundances of potassium and thorium on the surface of Mercury: Results from the MESSENGER gamma-ray spectrometer. *Journal of Geophysical Research* 117: E00L04. http://dx.doi.org/10.1029/2012JE004141.

Phillips RJ (1994) Estimating lithospheric properties at Atlas Regio, Venus. *Icarus* 112(1): 147–170.

Phillips RJ, Bullock MA, and Hauck SA (2001a) Climate and interior coupled evolution on Venus. *Geophysical Research Letters* 28: 1779–1782.

Phillips RJ, Zuber MT, Solomon SC, et al. (2001b) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291(5513): 2587–2591.

Phillips RJ, Zuber MT, Smrekar SE, Mellon MT, Head JW, Tanaka K, et al. (2008) Mars north polar deposits: Stratigraphy, age, and geodynamical response. *Science* 320(5880): 1182–1185.

Plesa AC and Breuer D (2014) Partial melting in one-plate planets: Implications for thermo-chemical and atmospheric evolution. *Planetary and Space Science*, special issue 'Planetary evolution and life', 98: 50–65, http://dx.doi.org/10.1016/j.pss.2013.10.007.

Plesa AC, Tosi N, and Breuer D (2014) Can a fractionally crystallized magma ocean explain the thermo-chemical evolution of Mars? *Earth and Planetary Science Letters* 403: 225–235. http://dx.doi.org/10.1016/j.epsl.2014.06.034.

Pozzo M, Davies C, Gubbins D, and Alfè D (2012) Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* 485: 355–358.

Pritchard ME and Stevenson DJ (1999) Thermal aspects of a lunar origin by giant impact. In: Canup R and Righter K (eds.) *The Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.

Prockter LM, Ernst CM, Denevi BW, et al. (2010) Evidence for young volcanism on Mercury from the third MESSENGER flyby. *Science* 329: 668–671. http://dx.doi.org/10.1126/science.1188186.

Purucker M, Ravat D, Frey H, Voorhies C, Sabaka T, and Acuña M (2000) An altitude-normalized magnetic map of Mars and its interpretation. *Geophysical Research Letters* 27: 2449–2452.

Ratcliff JT, Schubert G, and Zebib A (1996) Steady tetrahedral and cubic patterns of spherical shell convection with temperature-dependent viscosity. *Journal of Geophysical Research* 101: 25473–25484.

Ratcliff JT, Tackley PJ, Schubert G, and Zebib A (1997) Transitions in thermal convection with strongly variable viscosity. *Physics of the Earth and Planetary Interiors* 102: 201–212.

Reese CC, Solomatov VS, and Moresi L-N (1998) Heat transport efficiency for stagnant lid convection with dislocation viscosity: Application to Venus and Mars. *Journal of Geophysical Research* 103: 13643–13658.

Reese CC and Solomatov VS (2006) Fluid dynamics of local martian magma oceans. *Icarus* 184: 102–120.

Reese CC, Orth CP, and Solomatov VS (2010) Impact origin for the Martian crustal dichotomy: Half emptied or half filled?. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2009JE003506.

Reese CC, Solomatov VS, and Moresi L-N (1999) Non-Newtonian stagnant lid convection and magma resurfacing on Venus. *Icarus* 139: 67–80.

Richter FM (1978) Experiments on the stability of convection rolls in fluids whose viscosity depends on temperature. *Journal of Fluid Mechanics* 89: 553–560.

Riner MA, Bina CR, Robinson MS, and Desch JS (2008) Internal structure of Mercury: Implications of a molten core. *Journal of Geophysical Research* 113: E08013. http://dx.doi.org/10.1029/2007JE002993.

Ringwood AE (1975) Some aspects of the minor element chemistry of lunar mare basalts. *Moon* 12: 127–157.

Ringwood AE (1977) Composition of the core and implications for the origin of the earth. *Geochemical Journal* 11: 111–135.

Rivoldini A, van Hoolst T, and Verhoeven O (2009) The interior structure of Mercury and its core sulfur content. *Icarus* 201: 12–30. http://dx.doi.org/10.1016/j.icarus.2008.12.020.

Roberts GO (1979) Fast viscous Bernard convection. *Geophysical and Astrophysical Fluid Dynamics* 12: 235–272.

Roberts JH, Lillis RJ, and Manga M (2009) Giant impacts on early Mars and the cessation of the Martian dynamo. *Journal of Geophysical Research* 114: E04009. http://dx.doi.org/10.1029/2008JE003287.

Roberts JH and Zhong SJ (2006) Degree-1 convection in the Martian mantle and the origin of the hemisphere dichotomy. *Journal of Geophysical Research* 111(E6): E06013.

Rochette P, Fillion G, Ballou R, Brunet F, Oulladiaf B, and Hood L (2003) High pressure magnetic transition in pyrrhotite and impact demagnetization on Mars. *Geophysical Research Letters* 30: 1683. http://dx.doi.org/10.1029/2003GL017359.

Rochette P, Lorand J-P, Fillion G, and Sautter V (2001) Pyrrhotite and the remanent magnetization of SNC meteorites: A changing perspective on Martian magnetism. *Earth and Planetary Science Letters* 190(1–2): 1–12.

Rückriemen T, Breuer D, and Spohn T (2014) Key characteristics of the iron snow regime in Ganymede's core. In: 45th LPSC 2014, 17–21. March 2014, The Woodlands, Texas.

Ruedas T, Tackley PJ, and Solomon SC (2013a) Thermal and compositional evolution of the martian mantle: Effects of phase transitions and melting. *Physics of the Earth and Planetary Interiors* 216: 32–58. http://dx.doi.org/10.1016/j.pepi.2012.12.002.

Ruedas T, Tackley PJ, and Solomon SC (2013b) Thermal and compositional evolution of the martian mantle: Effects of water. *Physics of the Earth and Planetary Interiors* 220: 50–72. http://dx.doi.org/10.1016/j.pepi.2013.04.006.

Ruiz J, McGovern PJ, Jimenez-Diaz A, Lopez V, Williams JP, and Hahn BC (2011) The thermal evolution of Mars as constrained by paleo-heat flows. *Icarus* 215: 508–517.

Runcorn SK (1975) An ancient lunar magnetic dipole field. *Nature* 253: 701–703.

Runcorn SK (1977) Early melting of the Moon. In: *Proceedings of the Lunar Science Conference 8th*, Houston, TX, 14–18 March 1977, pp. 463–467.

Russell CT, Baker DN, and Slavin JA (1988) The magnetosphere of Mercury. In: Viulas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AZ: University of Arizona Press.

Russell CT, Elphic RC, and Slavin JA (1979a) Initial Pioneer Venus magnetic field results: Dayside observations. *Science* 203: 745.

Russell CT, Elphic RC, and Slavin JA (1979b) Initial Pioneer Venus magnetic field results: Nightside observations. *Science* 205: 114.

Saal AE, Hauri EH, Cascio ML, Van Orman JA, Rutherford MC, and Cooper RF (2008) Volatile content of lunar volcanic glasses and the presence of water in the Moon's interior. *Nature* 454: 192–195.

Samuel H and Tackley PJ (2008) Dynamics of core formation and equilibrium by negative diapirism. *Geochemistry, Geophysics, Geosystems* 9: Q06011. http://dx.doi.org/10.1029/2007GC001896.

Sandwell DT and Schubert G (1992) Evidence for retrograde lithospheric subduction on Venus. *Science* 257(5071): 766–770.

Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of the Earth and Planetary Interiors* 112: 43–54.

Sanloup C and Fei Y (2004) Closure of the Fe-S-Si liquid miscibility gap at high pressures. *Physics of the Earth and Planetary Interiors* 147: 57–74.

Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus: What are they telling us? *Journal of Geophysical Research* 97: 13257–13301.

Schenk P, Hargitai H, Wilson R, McEwen A, and Thomas P (2001) The mountains of Io: Global and geological perspectives from Voyager and Galileo. *Journal of Geophysical Research* 106: 33201–33222.

Schmeling H, Marquart G, and Ruedas T (2003) Pressure- and temperature-dependent thermal expansivity and the effect on mantle convection and surface observables. *Geophysical Journal International* 154: 2224–2229.

Schott B, van den Berg AP, and Yuen DA (2001) Focussed time-dependent Martian volcanism from chemical differentiation coupled with variable thermal conductivity. *Geophysical Research Letters* 28(22): 4271–4274.

Schubert G (1979) Subsolidus convection in the mantles of terrestrial planets. *Annual Review of Earth and Planetary Sciences* 7: 289–342.

Schubert G and Anderson CA (1985) Finite element calculations of very high Rayleigh number thermal convection. *Geophysical Journal of the Royal Astronomical Society* 80: 575–601.

Schubert G, Bercovici D, and Glatzmeier GA (1990) Mantle dynamics in Mars and Venus: Influence of an immobile lithosphere on three-dimensional mantle convection. *Journal of Geophysical Research* 95: 14105–14129.

Schubert G, Cassen P, and Young RE (1979a) Core cooling by subsolidus mantle convection. *Physics of the Earth and Planetary Interiors* 20: 194–208.

Schubert G, Cassen P, and Young RE (1979b) Subsolidus convective cooling histories of terrestrial planets. *Icarus* 38: 192–211.

Schubert G and Lingenfelter RE (1973) Martian center of mass – Center of figure offset. *Nature* 242: 251–252.

Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Viulas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AZ: University of Arizona Press.

Schubert G, Russell CT, and Moore WB (2000) Timing of the Martian dynamo. *Nature* 408: 666–667.

Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Jakosky BM, Snyder CW, Matthews MS, and Kieffer HH (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.

Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.

Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories, compositions, and internal structures of the moons of the solar system. In: Burns JA and Matthews M (eds.) *Satellites*, pp. 224–292. Tucson, AZ: University of Arizona Press.

Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press, 940 pp.

Schubert G, Turcotte DL, and Oxburgh ER (1970) Phase change instability in the mantle. *Science* 169: 1075–1077.

Schultz RA and Watters TR (2001) Forward mechanical modeling of the Amenthes Rupes thrust fault on Mars. *Geophysical Research Letters* 28: 4659–4662.

Schumacher S and Breuer D (2006) Influence of a variable thermal conductivity on the thermochemical evolution of Mars. *Journal of Geophysical Research* 111(E2), CiteID E02006.

Schutt DL and Lesher CE (2006) Effects of melt depletion on the density and seismic velocity of garnet and spinel lherzolite. *Journal of Geophysical Research* 111(B5), CiteID B05401.

Scott DR and Stevenson DJ (1986) Magma ascent by porous flow. *Journal of Geophysical Research* 91: 9283–9296.

Scott HP, Williams Q, and Ryerson FJ (2002) Experimental constraints on the chemical evolution of large icy satellites. *Earth and Planetary Science Letters* 203: 399–412.
Segatz M, Spohn T, Ross MN, and Schubert G (1988) Tidal dissipation, surface heat flow, and figure of viscoelastic models of Io. *Icarus* 75: 187–206.
Seipold U (1992) Depth dependence of thermal transport properties for typical crustal rocks. *Physics of the Earth and Planetary Interiors* 69(3–4): 299–303.
Senshu H, Kuramoto K, and Matsui T (2002) Thermal evolution of a growing Mars. *Journal of Geophysical Research* 107: 5118. http://dx.doi.org/10.1029/2001JE001819.
Sharpe HN and Peltier WR (1978) Parameterized mantle convection and the Earth's thermal history. *Geophysical Research Letters* 5: 737–740.
Sharpe HN and Strangway DW (1976) The magnetic field of Mercury and modes of thermal evolution. *Geophysical Research Letters* 3: 285–288.
Shea EK, Weiss BP, Cassata WS, et al. (2012) A long-lived lunar core dynamo. *Science* 335: 453–456. http://dx.doi.org/10.1126/science.1215359.
Shearer CK, Hess PC, Wieczorek MA, et al. (2006) Thermal and magmatic evolution of the moon. *Reviews in Mineralogy and Geochemistry* 60: 365–518.
Siegfried RW and Solomon SC (1974) Mercury – Internal structure and thermal evolution. *Icarus* 23: 192–205.
Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid: Topography admittance of Venus. *Science* 264: 798–803.
Simons M, Solomon SC, and Hager BH (1997) Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of Venus. *Geophysical Journal International* 131: 24–31.
Sleep NH (1994) Martian plate tectonics. *Journal of Geophysical Research* 99: 5639–5655.
Smith DE, Sjogren WL, Tyler GL, Balmino G, Lemoine FG, and Konopliv AS (1999) The gravity field of Mars: Results from Mars Global Surveyor. *Science* 286: 94–97.
Smith D, Zuber M, Phillips R, et al. (2012) Gravity field and internal structure of Mercury from MESSENGER. *Science* 336(6078): 214–217. http://dx.doi.org/10.1126/science.1218809.
Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implication for surface evolution. *Science* 284: 1495–1503.
Smrekar SE (1994) *Tectonics of Mantle Plumes on Venus*. Houston, TX: Lunar and Planetary Institute, p. 72, Abstract.
Smrekar SE and Sotin C (2012) Constraints on mantle plumes on Venus: Implications for volatile history. *Icarus* 217: 510–523.
Smrekar SE and Stofan ER (1999) Origin of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.
Smrekar SE, Stofan ER, Mueller N, et al. (2010) Recent hot-spot volcanism on Venus from VIRTIS emissivity data. *Science* 328(5978): 605–608.
Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110(E12), CiteID E12008.
Sohl F and Spohn T (1997) The structure of Mars: Implications from SNC-meteorites. *Journal of Geophysical Research* 102: 1613–1635.
Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity. *Physics of Fluids* 7: 266–274.
Solomatov VS and Moresi L-N (1996) Stagnant lid convection on Venus. *Journal of Geophysical Research* 101: 4737–4753.
Solomatov VS and Moresi L-M (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on Earth and other terrestrial planets. *Journal of Geophysical Research* 109: 21795–21817.
Solomatov VS and Stevenson DJ (1993) Nonfractional crystallization of a terrestrial magma ocean. *Journal of Geophysical Research* 98: 5391–5406.
Solomon SC (1976) Some aspects of core formation in Mercury. *Icarus* 28: 509–521.
Solomon SC (1977) The relationship between crustal tectonics and internal evolution in the Moon and Mercury. *Physics of the Earth and Planetary Interiors* 15: 135–145.
Solomon SC (1979) Formation, history and energetics of core in terrestrial planets. *Physics of the Earth and Planetary Interiors* 19: 168–182.
Solomon CS, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.
Solomon SC and Chaiken J (1976) Thermal expansion and thermal stress in the Moon and terrestrial planets: Clues to early thermal history. In: *Proceedings of the Lunar Science Conference*, vol. 7, pp. 3229–3243.
Solomon SC and Longhi J (1977) Magma oceanography: 1. Thermal evolution. In: *Proceedings of the 8th Lunar Science Conference*, Houston, TX, pp. 583–599.
Sotin C and Labrosse S (1999) Three-dimensional thermal convection in an iso-viscous, infinite Prandtl number fluid heated from within and from below: Applications to the transfer of heat through planetary mantles. *Physics of the Earth and Planetary Interiors* 112: 171–190.
Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–236.
Spohn T, Acuña MA, Breuer D, et al. (2001) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.
Spohn T and Breuer D (1993) Mantle differentiation through continental crust growth and recycling and the thermal evolution of the Earth. In: Takahashi E, Jeanloz R, and Rubie DC (eds.) *Evolution of the Earth and Planets*. *Geophysical Monograph Series 74*, vol. 14, pp. 55–71. Washington, DC: IUGG, American Geophysical Union.
Spohn T and Schubert G (1982) Modes of mantle convection and the removal of heat from the Earth's interior. *Journal of Geophysical Research: Solid Earth (1978–2012)* 87(B6): 4682–4696.
Spohn T and Schubert G (1983) Convective thinning of the lithosphere: A mechanism for continental rifting and mid-plate volcanism on Earth, Venus, and Mars. *Tectonophysics* 94: 67–90.
Spohn T, Sohl F, and Breuer D (1998) Mars. *Astronomy and Astrophysics Review* 8: 181–235.
Spudis PD and Guest JE (1988) Stratigraphy and geologic history of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 118–164. Tucson: University of Arizona Press.
Šrámek O and Zhong SJ (2012) Martian crustal dichotomy and Tharsis formation by partial melting coupled to early plume migration. *Journal of Geophysical Research* 117(E1): E01005. http://dx.doi.org/10.1029/2011JE003867.
Stacey FD and Anderson OL (2001) Electrical and thermal conductivities of Fe–Ni–Si alloy under core conditions. *Physics of the Earth and Planetary Interiors* 124(3): 153–162.
Stanley S, Bloxham J, Hutchinson WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.
Stanley S, Elkins-Tanton LT, Zuber MT, and Parmentier EM (2008) Mars' paleomagnetic field as the result of a single-hemisphere dynamo. *Science* 321: 1822–1825.
Stegman DR, Jellinek AM, Zatman SA, Baumgardner JR, and Richards MA (2003) An early lunar core dynamo driven by thermochemical mantle convection. *Nature* 421(6919): 143–146.
Steinbach V and Yuen DA (1992) The effects of multiple phase transitions on Venusian mantle convection. *Geophysical Research Letters* 19: 2243–2246.
Steinberger B, Werner SC, and Torsvik TH (2010) Deep versus shallow origin of gravity anomalies, topography and volcanism on Earth, Venus and Mars. *Icarus* 207(2): 564–577. http://dx.doi.org/10.1016/j.icarus.2009.12.025.
Stephenson A, Runcorn SK, and Collinson DW (1975) On changes in the intensity of the ancient lunar magnetic field. In: *Proceedings of the 6th Lunar Science Conference* (*Geochimica et Cosmochimica Acta*).
Stevenson DJ (1987) Origin of the Moon – The collision hypothesis. *Annual Reviews of Earth and Planetary Science* 15: 271–315.
Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 231–249. New York: Oxford University Press.
Stevenson DJ (2001) Mars core and magnetism. *Nature* 412: 214–219.
Stevenson DJ (2002) Possible connections between the history of the Venus magnetic field and observable features. *Spring Meeting 2002*, American Geophysical Union, Abstract #P21A-08.
Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
Stewart AJ, Schmidt MW, van Westrenen W, and Liebske C (2007) Mars: A new core-crystallization regime. *Science* 316(5829): 1323–1325.
Stofan ER, Bindschadler DL, Head JW, and Parmentier EM (1991) Corona structures on Venus – Models of origin. *Journal of Geophysical Research* 96(E4): 20933–20946.
Strom RG, Trask NJ, and Guest JE (1975) Tectonism and volcanism on Mercury. *Journal of Geophysical Research* 80: 2478–2507.
Suavet C, Weiss BP, Cassata WS, et al. (2013) Persistence and origin of the lunar core dynamo. *Proceedings of the National Academy of Sciences* 110: 8453–8458.
Snyder GA, Taylor LA, and Neal CR (1992) A chemical model for generating the sources of mare basalts: Combined equilibrium and fractional crystallization of the lunar magmasphere. *Geochimica et Cosmochimica Acta* 56: 3809–3823.
Tackley PJ (1996) On the ability of phase transitions and viscosity layering to induce long wavelength heterogeneity in the mantle. *Geophysical Research Letters* 23: 1985–1988.
Tackley PJ (2000) The quest for self-consistent incorporation of plate tectonics in mantle convection. In: Richards MA, Gordon R, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 121. Washington, DC: American Geophysical Union.
Tackley PJ, Schubert G, Glatzmeier GA, Schenk P, Ratcliff JT, and Matas JP (2001) Three-dimensional simulations of mantle convection in Io. *Icarus* 149: 79–93.

Tackley PJ, Stevenson DJ, Glatzmeier GA, and Schubert G (1993) Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the Earth's mantle. *Nature* 361: 699–704.

Tackley PJ, Stevenson DJ, Glatzmeier GA, and Schubert G (1994) Effects of multiple phase transitions in a three-dimensional spherical model of convection in Earth's mantle. *Journal of Geophysical Research* 99: 15877–15901.

Tan E, Choi E, Thoutireddy P, Gurnis M, and Aivazis M (2011) On the location of plumes and lateral movement of thermochemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12(7): Q07005.

Tanaka LK, Scott DH, and Greeley R (1992) Global stratigraphy. In: Jakosky BM, Snyder CW, Matthews MS, and Kieffer HH (eds.) *Mars*, pp. 345–382. Tucson, AZ: University of Arizona Press.

Tang X and Dong JJ (2010) Lattice thermal conductivity of MgO at conditions of Earth's interior. *Proceedings of the National Academy of Sciences* 107: 4539–4543. http://dx.doi.org/10.1073/pnas.0907194107.

Taylor LA, Shervais JW, Hunter RH, Shih D-Y, Bansal BM, and Nyquist LE (1983) Pre 4.2-Ga mare basalt volcanism in the lunar highlands. *Earth and Planetary Science Letters* 66: 33–47.

Tikoo SM, Weiss BP, Buz J, et al. (2012) Magnetic fidelity of lunar samples and implications for an ancient core dynamo. *Earth and Planetary Science Letters* 337: 93–103.

Toft PB and Arkani-Hamed J (1992) Magnetization of the Pacific Ocean lithosphere deduced from Magsat data. *Journal of Geophysical Research* 97: 4387–4406.

Toksöz MN, Hsui AT, and Johnston DH (1978) Thermal evolution of the Moon and the terrestrial planets. In: *The Soviet-American Conference on Cosmochemistry of the Moon and Planets*, pp. 245–328. NASA SP-370.

Tonks WB and Melosh HJ (1990) The physics of crystal settling and suspension in a turbulent magma ocean. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 151–174. New York: Oxford University Press.

Tonks WB and Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100: 326–346.

Tosi N, Grott M, Plesa A-C, and Breuer D (2013) Thermo-chemical evolution of Mercury's interior. *Journal of Geophysical Research: Planets* 118: 2474–2487. http://dx.doi.org/10.1002/jgre.20168.

Tosi N, Plesa A-C, and Breuer D (2013) Overturn and evolution of a crystallized magma ocean: A numerical parameter study for Mars. *Journal of Geophysical Research: Planets* 118(6): 1512–1528. http://dx.doi.org/10.1002/jgre.20109.

Tosi N, Yuen D, de Koker N, and Wentzcovitch R (2013) Mantle dynamics with pressure- and temperature-dependent thermal expansivity and conductivity. *Physics of the Earth and Planetary Interiors* 217: 48–58. http://dx.doi.org/10.1016/j.pepi.2013.02.004.

Tozer DC (1965) Heat transfer and convection currents. In: *A Symposium on Continental Drift. Philosophical Transactions of the Royal Society* 258: 252–271.

Trompert RA and Hansen U (1998) Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* 395: 686–689.

Tsuchiya T, Tsuchiya J, Umemoto K, and Wentzcovitch RM (2004) Phase transition in $MgSiO_3$ perovskite in the earth's lower mantle. *Earth and Planetary Science Letters* 224: 241–248.

Turcotte DL (1982) Thermal stresses in planetary elastic lithospheres. *Lunar and Planetary Science* XII: 809–810, Abstract.

Turcotte DL (1989) Crustal deformation: Earth VS Venus. In: *Abstracts for the Venus Geoscience Tutorial and Venus Geologic Mapping Workshop*, p. 49. Houston, TX: Lunar and Planetary Institute.

Turcotte DL (1993) An episodic hypothesis for Venusian tectonics. *Journal of Geophysical Research* 98: 17061–17068.

Turcotte DL and Huang J (1990) Implications on crust formation on Mars from parameterized convection calculations. In: *21st Lunar and Planetary Science Conference Abstracts*, 1266.

Turcotte DL and Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.

Turcotte DL and Schubert G (1982) *Geodynamics*. London: Wiley.

Usselman TM (1975) Experimental approach to the state of the core: Part I. The liquidus relations of the Fe-rich portion of the Fe–Ni–S system from 30 to 100 kb. *American Journal of Science* 275(3): 278–290.

Veeder GJ, Matson DL, Johnson TV, Davies AG, and Blaney DL (2004) The polar contribution to the heat flow of Io. *Icarus* 169(1): 264–270.

Vilim R, Stanley S, and Hauck SA (2010) Iron snow zones as a mechanism for generating Mercury's weak observed magnetic field. *Journal of Geophysical Research: Planets* 115(E11).

Wänke H and Dreibus G (1994) Chemistry and accretion history of Mars. *Philosophical Transactions of The Royal Society of London A* 349: 285–293.

Wark DA and Watson EB (1998) Grain-scale permeabilities of texturally equilibrated, monomineralic rocks. *Earth and Planetary Science Letters* 164: 591–605.

Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240.

Warren PH and Wasson JT (1979) The origin on KREEP. *Reviews of Geophysics* 17: 73–88.

Watters TR and Nimmo F (2010) The tectonics of Mercury. *Planetary Tectonics* 11: 15.

Watters TR, Robinson MS, Beyer RA, et al. (2010) Evidence of recent thrust faulting on the Moon revealed by the Lunar Reconnaissance Orbiter Camera. *Science* 329(5994): 936–940.

Watters TR, Robinson MS, Bina CR, and Spudis PD (2004) Thrust faults and the global contraction of Mercury. *Geophysical Research Letters* 31: 04701.

Watters TR, Robinson MS, and Cook AC (1998) Topography of lobate scarps on Mercury: New constraints on the planet's contraction. *Geology* 26: 991–994. http://dx.doi.org/10.1130/0091-7613/1998/026.

Watters T, Solomon S, Robinson M, et al. (2009) The tectonics of Mercury: The view after MESSENGER's first flyby. *Earth and Planetary Science Letters* 285: 283–296. http://dx.doi.org/10.1016/j.epsl.2009.01.025.

Weber RC, Lin PY, Garnero EJ, Williams Q, and Lognonne P (2011) Seismic detection of the lunar core. *Science* 331: 309–312. http://dx.doi.org/10.1126/science.1199375.

Weertman J and Weertman JR (1975) High temperature creep of rock and mantle viscosity. *Annual Review of Earth and Planetary Sciences* 3: 293–315.

Weidenschilling SJ (1978) Iron/silicate fractionation and the origin of Mercury. *Icarus* 35: 99–111.

Weinstein SA (1995) The effects of a deep mantle endothermic phase change on the structure of thermal convection in silicate planets. *Journal of Geophysical Research* 100(E6): 11719–11728.

Weinstein SA (1996) The potential role of non-Newtonian rheology in the resurfacing of Venus. *Geophysical Research Letters* 23: 511–514.

Weiss BP, Vali H, Baudenbacher FJ, Kirschvink JL, Stewart ST, and Shuster DL (2002) Records of an ancient Martian magnetic field in ALH84001. *Earth and Planetary Science Letters* 201(3–4): 449–463.

Wenzel MJ, Manga M, and Jellinek AM (2004) Tharsis as a consequence of Mars' dichotomy and layered mantle. *Geophysical Research Letters* 31(4), CiteID L04702.

Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.

Wetherill GW (1988) Accumulation of Mercury from planetesimals. In: Vilas F, et al. (ed.) *Mercury*, pp. 670–691. Tucson, AZ: University of Arizona Press.

Wieczorek MA (2008) Constraints on the composition of the Martian south polar cap from gravity and topography. *Icarus* 196: 506–517. http://dx.doi.org/10.1016/j.icarus.2007.10.026.

Wieczorek MA, Bradley LJ, Khan A, Pritchard ME, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364. http://dx.doi.org/10.2138/rmg.2006.60.3.

Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339: 671–675. http://dx.doi.org/10.1126/science.1231530.

Wieczorek MA and Phillips RJ (2000) The 'Procellarum KREEP Terrane': Implications for mare volcanism and lunar evolution. *Journal of Geophysical Research* 105: 20417–20430.

Wieczorek MA and Zuber MT (2004) Thickness of the Martian crust: Improved constraints from geoid-to-topography ratios. *Journal of Geophysical Research* 109: E01009. http://dx.doi.org/10.1029/2003JE002153.

Wilhelms DE and Squyres SW (1984) The Martian hemispheric dichotomy may be due to a giant impact. *Nature* 309: 138–140.

Williams Q (2009) Bottom-up versus top-down solidification of the cores of small solar system bodies: Constraints on paradoxical cores. *Earth and Planetary Science Letters* 284(3): 564–569.

Williams J-P and Nimmo F (2004) Thermal evolution of the Martian core: Implications for an early dynamo. *Geology* 32: 97–100.

Wise DU, Golombek MP, and McGill GE (1979) Tectonic evolution of Mars. *Journal of Geophysical Research* 84: 7934–7939.

Wolstencroft M, Davies JH, and Davies DR (2009) Nusselt–Rayleigh number scaling for spherical shell Earth mantle simulation up to a Rayleigh number of 109. *Physics of the Earth and Planetary Interiors* 176: 132–141.

Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites and a geophysical model of the Moon. In: *Proceedings of the Apollo 11 Lunar and Science Conference*, pp. 965–988.

Xu Y, Shankland TJ, Linhardt S, Rubie DC, Langenhorst F, and Klasinski K (2004) Thermal diffusivity and conductivity of olivine, wadsleyite and ringwoodite to 20 GPa and 1373 K. *Physics of the Earth and Planetary Interiors* 143–144: 321–336.

Yamashita N, Hasebe N, Reedy RC, et al. (2010) Uranium on the Moon: Global distribution and U/Th ratio. *Geophysical Research Letters* 37(10).

Yanagisawa T and Yamagishi Y (2005) Rayleigh-Bénard convection in spherical shell with infinite Prandtl number at high Rayleigh number. *Journal of the Earth Simulator* 4: 11–17.

Yoder CF (1979) How tidal heating in Io drives the Galilean orbital resonance locks. *Nature* 279: 767–770.

Yoder CF and Peale SJ (1981) The tides of Io. *Icarus* 47: 1–35.

Zaranek SE and Parmentier EM (2004) Convective cooling of an initially stably stratified fluid with temperature-dependent viscosity: Implications for the role of solid-state convection in planetary evolution. *Journal of Geophysical Research* 109: B03409. http://dx.doi.org/10.1029/2003JB002462.

Zebib A, Goyal AK, and Schubert G (1985) Convective motion in a spherical shell. *Journal of Fluid Mechanics* 152: 39–48.

Zebib A, Schubert G, Dein JL, and Paliwal RC (1983) Character and stability of axisymmetric thermal convection in spheres and spherical shells. *Geophysical and Astrophysical Fluid Dynamics* 23: 1–42.

Zhan X and Schubert G (2012) Powering Ganymede's dynamo. *Journal of Geophysical Research* 117(E8): E08011.

Zhang N, Parmentier EM, and Liang Y (2013) A 3D numerical study of the thermal evolution of the Moon after cumulate mantle overturn: The importance of rheology and core solidification. *Journal of Geophysical Research*. 118: 1789–1804. http://dx.doi.org/10.1029/jgre.20121.

Zhao Y-H, Zimmerman ME, and Kohlstedt DL (2009) Effect of iron content on the creep behavior of olivine: 1. Anhydrous conditions. *Earth and Planetary Science Letters* 287(1–2): 229–240.

Zhong S (2005) Dynamics of thermal plumes in three-dimensional isoviscous thermal convection. *Geophysical Journal International* 162: 289–300.

Zhong S (2009) Migration of Tharsis volcanism on Mars caused by differential rotation of the lithosphere. *Nature Geoscience* 2: 19–23. http://dx.doi.org/10.1038/NGEO392.

Zhong S, Parmentier EM, and Zuber MT (2000) A dynamic origin for the global asymmetry of lunar mare basalts. *Earth and Planetary Science Letters* 177: 131–140.

Zhong S and Zuber MT (2001) Degree-1 mantle convection and the crustal dichotomy on Mars. *Earth and Planetary Science Letters* 189: 75–84.

Zhu MH, Chang J, Ma T, et al. (2013) Potassium map from Chang'E-2 constraints the impact of Crisium and Orientale basin on the Moon. *Scientific Reports* 3: .

Ziethe R, Seiferlin K, and Hiesinger H (2009) Duration and extent of lunar volcanism: comparison of 3D convection models to mare basalt ages. *Planetary and Space Science* 57: 784–796. http://dx.doi.org/10.1016/j.pss.2009.02.002.

Zolotov M, Sprague A, Hauck S, Nittler L, Solomon S, and Weider S (2013) The redox state, FeO content, and origin of sulfur-rich magmas on Mercury. *Journal of Geophysical Research* (118). http://dx.doi.org/10.1029/2012JE4274.

Zuber MT, Solomon SC, Phillips RJ, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

This page intentionally left blank

10.09 Planetary Tectonics and Volcanism: The Inner Solar System

TKP Gregg, University at Buffalo, Buffalo, NY, USA

© 2015 Elsevier B.V. All rights reserved.

10.09.1	**Introduction**	307
10.09.2	**Planetary Volcanism and Tectonics**	307
10.09.2.1	Tectonics	307
10.09.2.2	Planetary Lithospheres	307
10.09.2.3	Mare-Type Wrinkle Ridges	308
10.09.3	**The Moon**	309
10.09.3.1	Lunar Tectonism	309
10.09.3.2	Lunar Volcanism	310
10.09.4	**Mars**	311
10.09.4.1	Martian Tectonics	311
10.09.4.2	Mars' Volcanism	313
10.09.5	**Venus**	314
10.09.5.1	Venus Tectonism	315
10.09.5.2	Volcanism on Venus	316
10.09.6	**Mercury**	318
10.09.6.1	Mercury Tectonics	318
10.09.6.2	Mercury Volcanism	320
10.09.7	**Summary**	321
References		321

10.09.1 Introduction

The Earth is but one of the many planetary bodies in the solar system (in this chapter, a 'planetary body' encompasses any solid body orbiting the Sun that is larger than a few hundred kilometers in diameter) and therefore represents only one datum. It is essential to examine geologic processes on other planetary bodies to gain the most complete understanding. If a model works well on Earth, a vital test for the validity of that model is to determine if it works equally well when applied to places in the solar system with different ambient conditions (e.g., gravity, surface pressure, and atmospheric composition). Thus, the investigation of volcanic and tectonic processes on other planetary bodies provides insight into those same processes operating on Earth.

In the last 20 years, terabytes of data have poured down to Earth from spacecraft orbiting Venus, Mars, the Moon, Saturn, Jupiter, and even a few asteroids. Information gleaned from these missions has revealed extraterrestrial processes that in some cases are quite familiar and, in others, remain poorly understood. In this chapter, the current understanding of volcanic and tectonic processes operating in the inner solar system (Mercury, Venus, the Earth/Moon, and Mars) are summarized. A single chapter cannot possibly cover all the information available – the entire outer solar system is avoided here – and therefore, the reader is encouraged to pursue additional readings (see, e.g., Watters and Schultz, 2010).

10.09.2 Planetary Volcanism and Tectonics

10.09.2.1 Tectonics

Both volcanic and tectonic processes are controlled largely by lithospheric and asthenospheric behaviors on Earth; other planetary bodies large enough to experience differentiation (larger than a few hundred kilometers in radius) should also have (or have had in the past) similar mechanical layering. Earth's tectonics are dominated by the lateral movement of rigid lithospheric plates over the weaker asthenosphere. This style of tectonism – plate tectonics – is unique to Earth and localizes tectonic activity primarily to plate boundaries. The other terrestrial planets (see Table 1) display movement and deformation of a single, continuous lithospheric shell. On the Moon, tectonics are locally controlled by large ($>10^2$ km radius) impact basins. Mars' tectonics are regionally controlled by both large impact basins and volcanic provinces (such as the Tharsis region). Large impact basins have tectonic signatures on Mercury, but there are also more global patterns of tectonism that are unique among the terrestrial planets. Venusian tectonism is expressed almost globally across the Venusian plains. Locally, quasicircular tectonic terrains called 'coronae' (up to thousands of kilometers across) and rift valleys called 'chasmata' reveal lithospheric extension and compression.

The icy satellites of Jupiter and Saturn also clearly display tectonic features, but there are fewer data available to constrain the mechanical layering inside those planetary bodies. Observations to date support that most of these icy satellites are also 'one-plate planets' whose tectonics are controlled mostly by large impact basins and tidal forces, with volcanic processes exhibiting local dominance.

10.09.2.2 Planetary Lithospheres

Tectonics are the movements of the lithosphere, so to understand planetary tectonics, it is first necessary to learn about the properties of these extraterrestrial lithospheres and the limitations of that knowledge (see Chapter 10.05, and references

Table 1 Physical parameters of the terrestrial planets

Planet	*Equatorial radius (km)*	*Bulk density (kg m^{-3})*	*Surface gravity (m s^{-2})*	*T$_E$*[a]	*Atmospheric pressure*[b] *(Pa)*
Mercury	2439	5430	2.78	172	0
Venus	6051	5250	8.60	22	9.12×10^6
Earth	6378	5520	9.78	34	1.01×10^6
Moon	1738	3340	1.62	103	0
Mars	3393	3950	3.72	126	600

[a]From Melosh (2011), Table 4.1.
[b]Value given for sea level (Earth) or mean planetary radius.

therein). Earth's internal layering, and the geophysical properties of those layers, has been revealed through painstaking collection and analyses of seismic data. As every geology initiate knows, a minimum of three seismic stations are required to locate the epicenter of an earthquake. Neither Mercury nor Venus has supported seismometers. Although both Viking Landers 1 and 2 contained seismometers, only Viking Lander 2 returned seismic data from Mars. Apollo missions 12, 14, 15, and 16 successfully deployed seismometers on the lunar surface (the Apollo Passive Seismic Experiment, or PSE) that continued to radio data to Earth until they were switched off in 1977 (cf. Nakamura, 1980, 2003, 2005). All of these seismometers were necessarily deployed on the lunar nearside, providing suboptimal data collection. Recent reprocessing of the PSE data, however, reveals the vital information about the structure and composition of the lunar interior (Weber et al., 2011; **Chapter 10.03**): the Moon likely contains a solid inner core, a liquid outer core, and a partially molten boundary layer (similar to Earth's internal structure). These new insights using old data are granted by improved technologies and suggest that we can look forward to improved understanding of deep planetary interiors.

Information about extraterrestrial lithospheres, and their elastic thicknesses (T_E), has come primarily from topography and gravity models (Belleguic et al., 2005; Grott and Breuer, 2009; Kiefer and Li, 2009; McGovern et al., 2002; Wieczorek, 2007; Wieczorek et al., 2006, 2013; **Chapter 10.05**). The lithosphere is defined as the outermost mechanical (as opposed to chemical) layer of a planetary body that behaves rigidly over geologic timescales. The effective lithospheric elastic thickness characterizes the apparent strength of the lithosphere. On Earth, the value of T_E and the depth to the base of the mechanical lithosphere are approximately the same for oceanic lithosphere, but can differ widely for continental lithosphere (e.g., Burov and Diament, 1995). Most simply, T_E is controlled by lithospheric composition, surface temperature, and heat flow (Melosh, 2011):

$$T_E = \frac{k\left(\frac{T_m}{2} - T_s\right)}{q_s} \quad [1]$$

where k is the thermal conductivity (J s^{-1} K^{-1}), T_s is the surface temperature of the planet, T_m is the melting temperature of the lithosphere, and q_s is the heat flow from the planet's surface (J s^{-1} m^{-2}). Thermal conductivity does not have a wide range among geologic materials (e.g., Navrotsky, 1995). Using this equation, and reasonable geologic assumptions, gives T_E of ~34 km for Earth; ~100 km for the Moon and Mars; ~20 km for Venus; and almost 200 km for Mercury (Melosh, 2011). These numbers predict distinct responses to volcanic loading or upwelling on each planet, for example.

It is important to note, however, that the values for T_E calculated by eqn [1] can give quite different values from those calculated using other methods. For example, Zuber et al. (2010) estimated T_E for Mercury to be on the order of 30–60 km, based on topographic measurements of contractional features. Smith et al. (2012) used orbital data collected from the MESSENGER (MErcury Surface, Space ENvironment, GEochemistry, and Ranging) spacecraft to estimate Mercury's T_E to be only 70–90 km. These variations reflect a combination of data resolution and the complexities (such as lithospheres composed of layered materials with different rheologies) inherent in the parameters controlling T_E.

Lithospheric or crustal deformation is responsible for the tectonic features observed on other planetary bodies. The crust, a portion of the lithosphere, is distinguished from the lithosphere by composition: in differentiated bodies, the crust is the outermost layer, distinguished by having the lowest density. On Earth, the crust is the outermost portion of the lithosphere, and it is likely that similar arrangements exist on other terrestrial planets. The large density contrast between the crust and mantle makes it somewhat simpler to model the crustal thickness than the lithospheric thickness.

Modeling the thickness of the crust and lithosphere requires detailed measurements of the surface topography and gravity. These data sets currently exist for Mars and Venus; MESSENGER data from Mercury and GRAIL (Gravity Recovery and Interior Laboratory) data for the Moon (see **Table 2**) are improving our modeling capabilities for those planetary bodies. Mars' crust displays a strong dichotomy: in the southern highlands, the crust is ~60 km thick – about twice as thick as in the northern lowlands (Neumann et al., 2004). On Venus, the spatial resolution of the gravity and topographic data provide fewer model constraints; recent work (James et al., 2013) suggests that the mean crustal thickness on Venus is 8–25 km. Data from GRAIL indicate a lunar crustal thickness between 34 and 43 km (Wieczorek et al., 2013). Acknowledging that the available data are relatively poor quality for the southern hemisphere of Mercury, Smith et al. (2012) calculated crustal thicknesses ranging from ~20 to 80 km, with a mean of 50 km.

10.09.2.3 Mare-Type Wrinkle Ridges

Mare-type wrinkle ridges (or, simply, 'wrinkle ridges') appear on every terrestrial planet (Watters, 1988), but are relatively rare on Earth (**Figure 1**). These features were first described on the lunar surface, where they are linear to arcuate in planform and have a superficial resemblance to folds (or wrinkles) in a rug – hence their name. Complete wrinkle ridges consist of a broad topographic arch topped by a narrow sinuous ridge (Golombek et al., 2001; Watters, 1988); locally, the sinuous ridge may meander off of the broad arch. They are asymmetrical in a topographic profile drawn perpendicular to their long axis, and the sense of asymmetry may change along ridge length. Wrinkle ridges are now recognized to be blind (not intersecting the surface) thrust faults (Golombek et al., 2001;

Table 2 Missions measuring the highest-resolution topography of the terrestrial planets

Planet	*Mission*	*Instrument*	*Year*	*Topographic resolution (m): footprint size*	*Relative vertical accuracy (m)*
Mercury	MESSENGER	Mercury Laser Altimeter (MLA)	2008 to present	15–100 m	0.10
Venus	Magellan	Altimeter	1990–93	8000–27 000	80
Moon	Lunar Reconnaissance Orbiter	Lunar Orbiter Laser Altimeter (LOLA)	2009 to present	150	0.10
Mars	Mars Global Surveyor	Mars Orbiter Laser Altimeter (MOLA)	1997–2001	168	1

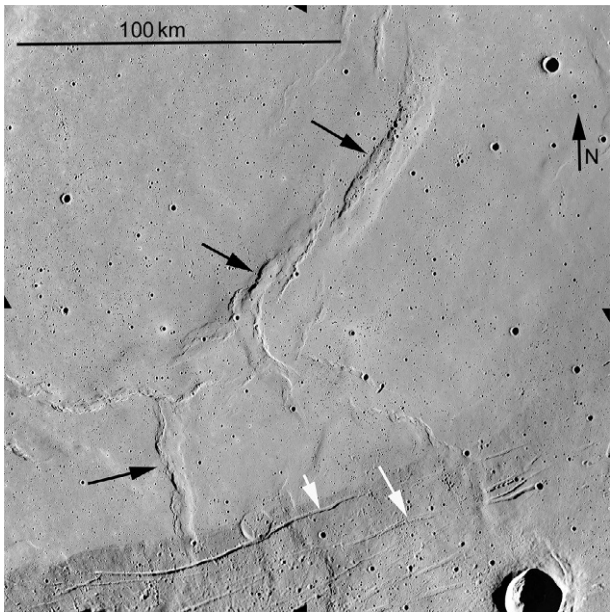


Figure 1 Mare-type wrinkle ridges (black arrows) are the most common tectonic feature in the inner solar system. Apollo 17 metric frame showing southern Mare Serenitatis (Apollo Image AS17-M-0451 (NASA/JSC/Arizona State University)). Black arrows point to mare-type wrinkle ridges; note broad arch and superposed sinuous ridge. White arrows point to troughs, interpreted to be graben. Image centered at 19.9° N, 24.1° E.

Schultz, 2000) through layered materials, with surface folding, although the precise nature of the faulting remains unclear (Golombek et al., 2001; Schultz, 2000; Watters, 2004; Zuber, 1995). The size and spacing of parallel wrinkle ridges can be used to determine the total thickness of the deformed material (e.g., Watters, 1993; Zuber, 1995): large spacing (10–10^2 km) reflects a greater thickness of deformed material than does small spacing (10s of kilometers) (Montesi and Zuber, 2003a,b; Watters, 2004; Zuber and Aist, 1990).

10.09.3 The Moon

10.09.3.1 Lunar Tectonism

Lunar tectonics are dominated by large (>10^2 km across) impact basins (Watters and Johnson, 2010; Wilhelms, 1987). Wrinkle ridges and linear or arcuate troughs are commonly found within or adjacent to impact basins (Watters and Johnson, 2010). Wrinkle ridges and lobate scarps are compressional features; linear or arcuate troughs are generally interpreted to be graben and therefore extensional. Lobate scarps are more broadly distributed across the lunar surface (Watters and Johnson, 2010) and are interpreted to be thrust faults (Binder, 1982; Watters et al., 2010).

Lunar impact basins on the nearside are filled with various amounts of basalts; these lowlands are referred to as maria (mare = singular). Wrinkle ridges are found within the maria and are typically both radial and concentric to the basin center (Maxwell et al., 1975; Watters and Johnson, 2010). Wilhelms (1987) pointed to wrinkle ridge rings within large lunar basins as marking the location of buried basin ring complexes, suggesting that premare topography may play an important role in the spatial distribution of lunar wrinkle ridges. Alternatively, lithospheric loading by basaltic lava may have caused basin subsidence, resulting in concentric and radial compressional stresses (Freed et al., 2001; Maxwell et al., 1975; Solomon and Head, 1979).

Linear or arcuate troughs (graben) are also associated with large lunar impact basins and their maria. They are interpreted to be graben because of their flat floors, inwardly dipping walls, and symmetrical topographic cross sections (**Figure 1**); debate remains about the precise nature of these graben: they may be the expression of near-surface dikes (Head and Wilson, 1993) or have been formed by localized basin stresses (Golombek, 1979). These graben are typically located entirely within maria deposits (and therefore are likely caused by basin-controlled extension) but locally may extend for a few kilometers into the adjacent highlands. No graben are observed on the lunar farside (Watters and Johnson, 2010) until recently: the Lunar Reconnaissance Orbiter Camera (LROC) revealed small-scale (locally as shallow as ~1 m) graben in the farside highlands and mare basalts (Watters et al., 2012a). These graben may be as young as 50 My (Watters et al., 2012a).

A unique class of impact crater, floor-fractured craters display a complex, spider-web-like network of graben on their otherwise flat floors (**Figure 2**; Wichman and Schultz, 1995; Wilhelms, 1987). Alphonsus crater (2.1° E, 13.5° S) is a type example. The current consensus is that these types of graben networks are formed by localized uplift of the crater floor, although the specific cause of the uplift remains unclear. Results of numerical modeling (Dombard and Gillis, 2001) suggest that viscous relaxation of the crater floor is unlikely to cause floor fractures in craters with diameters < 60 km. Uplift caused by a local igneous intrusion is the most likely

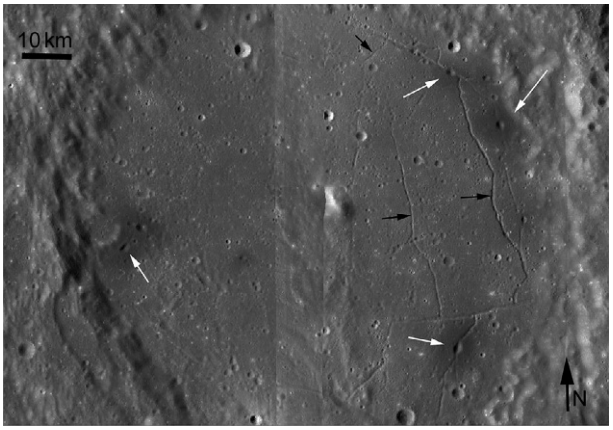


Figure 2 Alphonsus crater (2.1° E, 13.5° S) on the Moon. Black arrows point to troughs (graben) on the crater floor; white arrows point to dark mantling deposits (pyroclastic deposits). Image from Lunar Reconnaissance Orbiter Wide-Angle Camera, courtesy of NASA/JSC/ASU.

explanation for floor fracturing of larger impact craters (Wichman and Schultz, 1995).

The spatial correlation of wrinkle ridges and linear or arcuate graben with lunar maria strongly suggests a causal (and therefore temporal) relationship. Certainly, many of the largest lunar impact basins on the nearside are associated with mass concentrations ('mascons') that could also be related to deformation of basin-weakened lithosphere (Solomon and Head, 1979, 1980). Crosscutting relations and impact crater size–frequency distributions suggest that both wrinkle ridges and graben formed shortly after the maria were deposited. The detailed examination of maria deposits cut by lunar graben, combined with impact crater size–frequency distributions, suggests that no maria younger than 3.6 Ga (±0.2 Ga) are crosscut by linear or arcuate graben (Lucchitta and Watkins, 1978). In contrast, wrinkle ridges are observed to deform the oldest (~4.0 Ga) and the youngest (1.2 Ga) maria (Hiesinger et al., 2003).

Lobate scarps (interpreted to be thrust faults (Schultz, 1976)) have been estimated to be even younger than the youngest wrinkle ridges, based on their relatively fresh appearance (**Figure 3**) and impact crater size–frequency distributions (Binder and Gunga, 1985). It has been proposed that some lunar lobate scarps may even currently be active (Watters et al., 2010, 2012a).

10.09.3.2 Lunar Volcanism

Lunar volcanism is concentrated on the lunar nearside within impact basins. These maria cover $\sim 6.3 \times 10^6$ km^2 (Head, 1976) with a volume of $\sim 5.0 \times 10^6$ km^3 (Budney and Lucey, 1998; Hiesinger et al., 2002; Horz, 1978). The Apollo astronauts returned over 350 kg of lunar material (NASA, 2007), and it is primarily from these samples (as well as rare meteorites from the Moon that have struck the Earth) that we know the composition of lunar lavas. By using these lunar samples as 'ground truth' for comparison with satellite data, the composition of the lunar surface is fairly well constrained.

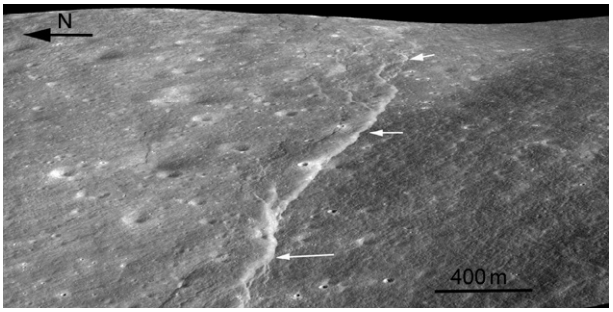


Figure 3 Lobate scarp (white arrows) within Slipher crater (160.1° E, 48.4° N) imaged with LROC narrow-angle camera. Scale bar is accurate in the foreground only; image depth is 3.5 km. Image courtesy NASA/GSFC/ASU (see http://wms.lroc.asu.edu/lroc_browse/view/M123514622).

Table 3 Lunar and martian lava compositions. All are consistent with basaltic (mafic) or ultramafic compositions

Oxide	*The Moon: Apollo 11, hi-Ti, low-K (#10020)*[a]	*Nakhla meteorite (Mars)*[b]	*Venus*[c] *Venera 14 Landing site*	*Earth basalt*[d]
SiO_2	39.79	48.6	50.0	49.97
TiO_2	10.62	0.34	1.28	1.87
Al_2O_3	9.93	1.68	18.4	15.99
Fe_2O_3				3.85
FeO	19.21	20.6	9.0	7.24
MnO	0.26	0.49	0.16	0.20
MgO	7.93	12.1	8.3	6.84
CaO	11.32	14.7	10.6	9.62
Na_2O	0.39	0.46	2.0	2.96
K_2O	0.04	0.13	0.21	1.12
Cr_2O_3	0.39	0.26		
P_2O_5				0.35
S		0.026		
H_2O		0.057		
SUM	99.88	99.6		
Mg/ Mg + Fe	0.35			
CaO/Al_2O_3	1.14			

[a]BVSP (Basaltic Volcanism Study Project) (1981).
[b]Lodders (1998).
[c]Kargel et al. (1993).
[d]Best and Christiansen (2001).

The lunar maria are 'basalts,' but these are unlike any basalts observed on Earth, and there is a range in lunar basalt compositions (see **Table 3**; Shearer et al., 2006). Although both lunar and terrestrial basalts share a relatively low SiO_2 content (<54 wt%), the abundances of the remaining oxides belie this similarity. Lunar basalts have more iron and less aluminum than typical terrestrial basalts, giving them a high melting temperature (~1400 °C) and a low viscosity (<1 Pa s at liquidus temperature) (Murase and McBirney, 1970). Furthermore, H_2O and CO_2 are the most abundant volcanic volatiles on Earth (e.g., Schmincke, 2004); in the anhydrous and reducing environment of the Moon, CO is the most probable volatile (Wilson and Head, 1981). In addition to these

compositional differences, the lunar volcanic environment has a lower gravity, no atmosphere, and a surface temperature that ranges from 126 to 373 K (see Wilson and Head, 1981). It should not be surprising, therefore, that lunar volcanism displays different morphologies than we see for basaltic volcanism on Earth.

Lunar basalts are concentrated in large impact basins on the nearside, although they also exist as lava 'ponds' (<2000 km^2) (Yingst and Head, 1997). Maria that are obscured by superposed, higher albedo material (such as impact crater ejecta) are termed 'cryptomaria' (Antonenko et al., 1994). The stratigraphic position of cryptomaria suggests that it represents some of the oldest volcanism on the Moon. However, this same stratigraphic position makes it difficult to identify and study.

The low viscosity of lunar basalts makes the preservation of lava flow margins rare, although some fantastically large (>1200 km long) examples can be seen in the Imbrium basin (Schaber, 1973; Schaber et al., 1976). Furthermore, low-viscosity lavas are unlikely to pile up around the vent to create tall volcanic constructs. Finally, low-viscosity lavas are likely to flood low-lying terrain, potentially burying their own vents. Thus, although lunar lavas comprise 17% of the lunar nearside, vents and edifices are relatively rare and tend to be concentrated in a few geographic locations: the Marius Hills region, the Aristarchus plateau region, and the Gruithuisen domes.

The Marius Hills region (**Figure 4**) is characterized by mounds and hills that have been interpreted to be volcanic domes and cones (Weitz and Head, 1999). Domes have flank slopes of 2° and cones are steeper; both have basal diameters <25 km. All currently available data are consistent with these hills being basalt (rather than a more evolved, silica-rich, viscous lava). The analysis of recently acquired topographic data from the Lunar Orbiter Laser Altimeter led Spudis (2011) to suggest that the entire Marius Hills complex is a large (330 km across and 3.2 km tall) shield volcano peppered with smaller parasitic vents. Lawrence et al. (2013) stated that the Marius Hills represent eruptions of mare materials and the morphological variability was caused by a range of eruption and emplacement conditions.

Figure 4 Marius Hills region on the Moon (centered at 14° N, 56° W). Note the volcanic mounds and the sinuous rilles; north is at the top. Marius Hills low-sun controlled NAC mosaic (20 m per pixel). Image courtesy of NASA/JPL/ASU.

Surrounding the Marius Hills is a high spatial density of sinuous rilles (**Figure 4**). Sinuous rilles are long (up to hundreds of kilometers) sinuous troughs that maintain a relatively constant width (generally <5 km) along their length. These troughs are generally u-shaped in cross section (as opposed to v-shaped) and some of them originate in circular or irregular depressions (Greeley, 1971; Schubert et al., 1970). These rilles formed via lava, although there remains debate as to whether they are erosional (lava downcutting into the preexisting surface by melting and assimilating the country rock) or constructional (e.g., lava tubes whose roofs have collapsed over time or drained lava channels) features (e.g., Williams et al., 2000 and references therein). There are hundreds of sinuous rilles on the lunar surface, and it is unlikely that all of them formed in precisely the same way: it is probable that some combination of constructional and erosional processes is responsible for their formation (Roberts, 2013).

The low viscosity of lunar basalts and lack of water would predict an abundance of effusive volcanism (i.e., lava flows) rather than explosive volcanism. However, astronaut Jack Schmidt returned a sample of orange soil from Apollo 17; analyses revealed that the soil contained microscopic droplets of orange volcanic glass. Subsequently, green glass droplets were identified in the Apollo 15 samples. These droplets formed by explosive eruptions of lunar basalts (Meyer, 2005): in the absence of atmospheric drag, the magma fragmented by bubbles (probably of CO) popping on the lunar surface formed microscopic beads. Dark mantle deposits (DMDs) are easily observed on the floor of Alphonsus crater (**Figure 3**), and the spatial association of these DMDs with small craters and the floor fractures (interpreted to be graben) indicates that the DMDs are pyroclastic deposits (Weitz et al., 1998). DMDs are now recognized across the lunar maria (Weitz et al., 1998), representing explosive lunar eruptions.

The timing of lunar volcanism is determined through a combination of (1) radiometric dating of Apollo-returned samples, (2) careful geologic mapping to determine maria stratigraphy, and (3) impact crater size–frequency distributions. Radiometric dating of Apollo-returned samples gives ages of 3.6–3.9 Ga (Apollo 11 and Apollo 17 high-titanium basalts) to 3.16–3.4 Ga (Apollo 12 and 15 low-titanium basalts) (Nyquist and Shih, 1992; Snyder et al., 2000). Mapping and impact crater size–frequency distributions suggest that the youngest maria patches on the lunar nearside may be only 1.0–1.2 Ga (Hiesinger et al., 2003; Schultz and Spudis, 1983). Lava production on the Moon appears to have peaked 3.5–3.8 Ga (Shearer et al., 2006), indicating that the Moon is a volcanically 'old' planet. However, there is evidence that Ina, a D-shaped crater about a kilometer across (18.6° N, 5.3° E), may still be actively degassing (Schultz et al., 2006) although the amount and type of volatiles remain unknown.

10.09.4 Mars

10.09.4.1 Martian Tectonics

Mars, like the Moon, is a one-plate planet (**Figure 5**), but can be divided into three distinct physiographic, topographic, and geologic provinces: (1) the northern lowlands, (2) the southern highlands, and (3) the Tharsis region (see Golombek and

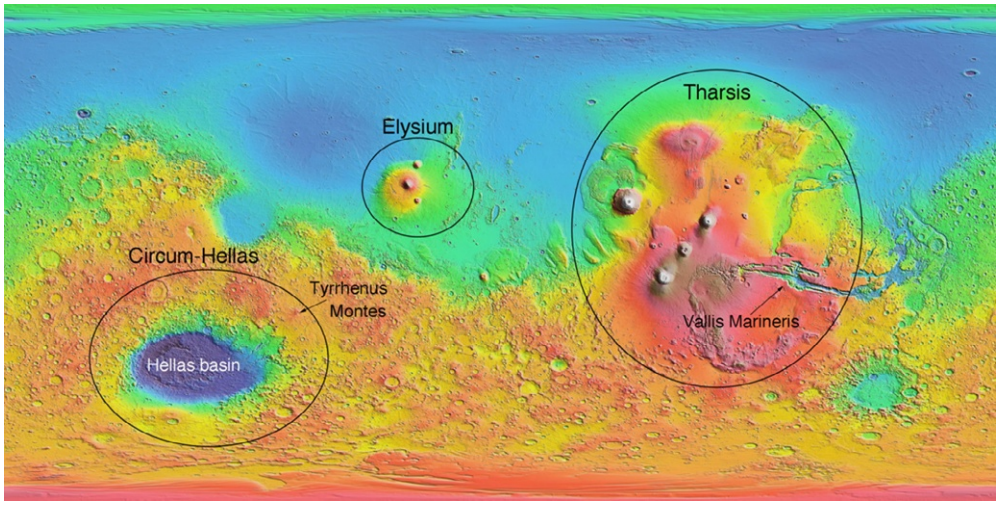


Figure 5 Topographic map of Mars: blues and purples are low elevations; reds and whites are high elevations. The whites in the Tharsis region are ~25 km above mean planetary radius; the purples in Hellas basin are about 12 km below mean planetary radius. Image courtesy of JPL/NASA.

Phillips, 2010 and references therein). The northern hemisphere is topographically low and relatively smooth, whereas the southern hemisphere is rugged with impact craters and is topographically higher (McCauley et al., 1972); this is commonly referred to as the 'crustal dichotomy.' Locally, the dichotomy boundary is marked by a scarp that may be as tall as 5.5 km (Aharonson et al., 2001; Watters et al., 2007); elsewhere, volcanics and erosional processes have modified (or buried) the nature of the boundary. There is evidence for some movement along this scarp in the form of compressional (lobate) scarps, extensional features such as graben, and broad topographic signatures of lithospheric flexure (Watters, 2003a,b; Watters et al., 2007). The similar ages for terrains north and south of the dichotomy boundary (e.g., Frey, 2006; Watters et al., 2007) argue against a plate-tectonic genesis for the dichotomy boundary. The origin of this dichotomy remains a mystery (e.g., Citron and Zhong, 2012), but it appeared early (>3.6 Ga) in Mars' history (Frey, 2006; Frey et al., 2002; Hartmann, 2005; Hartmann and Neukum, 2001; Nimmo et al., 2008; Watters et al., 2007).

Although ancient seafloor spreading has been proposed as a possible origin for the crustal dichotomy (Sleep, 1994), there is no morphological or chronological evidence for plate tectonics at the dichotomy boundary (e.g., Frey, 2006; Pruis and Tanaka, 1995). However, there is an interesting magnetic signature in the oldest portions of the southern highlands (Acuna et al., 2001). The alternating bands of strong/weak magnetic fields in the southern highlands are reminiscent of the alternating magnetic bands found in the ocean crust around Earth's mid-ocean ridges and require that Mars once generated its own active global magnetic field. The location of the magnetic pattern requires that Mars' dynamo stopped working within a few hundred million years of planetary formation (Acuna et al., 2001; Hartmann and Neukum, 2001), leaving behind no geologic signature of rifting associated with the magnetic stripes.

The Tharsis region is the largest volcanic province on Mars, and its tectonic influence spans over half of the planet. Volcanic products (primarily lava flows) constructed a pile ~5000 km across and 9 km tall during the first billion or so years of Mars' history (Carr and Head, 2010; Phillips et al., 2001); currently, the volcanic pile is approximately 10 km high. The enormous mass of this volcanic pile deformed the martian lithosphere, creating a circumferential trough, an antipodal high, and gravity anomalies (Phillips et al., 2001). Circumferential wrinkle ridges surrounding the Tharsis region (indicating compression) are particularly visible in Lunae Planum (Watters, 1988, 1993; Zuber and Aist, 1990), where their regular spacing (~60 km; Zuber, 1995) suggests that the thickness of the deformed layer is also ~60 km (a 'thick-skinned' interpretation). Alternatively, the wrinkle ridges may be the result of deformation of only the volcanic plains (a few kilometers thick – a 'thin-skinned' interpretation) (Watters, 2004). Most striking is the largest canyon in the solar system, Vallis Marineris (4000 km long, up to 7 km deep, and 200 km wide) (**Figure 5**). The parallel, inward-dipping walls and flat floor of this canyon are consistent with Vallis Marineris being a large rift valley system. Geophysical modeling results show that the distribution and orientation of most tectonic features around the Tharsis region can be explained by flexural loading of the lithosphere, using present-day values for gravity and topography (Golombek and Phillips, 2010). These results, combined with the timing of the deformation, indicate that the martian lithosphere beneath the Tharsis region has remained unchanged since about 3.6 Ga. Recent modeling by Beuthe et al. (2012) suggests that the elastic lithosphere thickness varies throughout the Tharsis region, being thickest beneath Olympus Mons and thinner elsewhere.

However, the thickness of Mars' elastic lithosphere is not everywhere the same nor was it likely the same at every stage in Mars' history. Published elastic lithosphere thickness estimates range from a mere 2 km thick beneath portions of Ascraeus Mons in the most recent period of Mars' history to possibly more than 300 km thick beneath Apollinaris Mons during the same time (see Golombek and Phillips, 2010, and references therein).

Mars displays a staggering variety of tectonic features: wrinkle ridges (thrust faults) and linear to arcuate troughs (graben) are abundant, but strike-slip faults are rare (Andrews-Hanna et al.,

2008; Schultz, 1989; Watters, 1992). Those documented to date are primarily accommodation features, marking regions where the regional stress field experiences a change in orientation (Andrews-Hanna et al., 2008). Andrews-Hanna et al. (2008) mapped two populations of strike-slip fault west of the Tharsis region, with individual offsets as long as 9 km. Stratigraphic relations suggest that these faults were active over a roughly 2 Ga period (from >3.8 to <2.8 Ga). Strike-slip faults located northwest of the Tharsis region are younger, active within the last ~2.0 Ga of Mars' history. The shift in space and time of strike-slip faulting around the Tharsis region requires a changing stress field around the Tharsis region with time.

Martian tectonic features are concentrated around volcanic provinces and large impact basins. As technology and image resolution improve, smaller and smaller tectonic features can be observed, but the same general patterns are still evident.

Martian wrinkle ridges are morphologically similar to their lunar counterparts (**Figure 6**) (Schultz and Watters, 2001; Watters, 1993, 2003a). They are most commonly found deforming plains materials that were emplaced around 3.2–3.6 Ga (Hartmann and Neukum, 2001; Hartmann, 2005). Near the Tharsis region (e.g., in Lunae Planum), wrinkle ridges are circumferential to the approximate center of the Tharsis region and their topography steps downward away from the Tharsis region, consistent with a thrust fault origin (Watters, 1993). Away from the Tharsis region, in Hesperia Planum, wrinkle ridges form roughly orthogonal, crosscutting networks that are oriented radial and circumferential to Hellas basin. Goudy et al. (2005) determined that the radially oriented ridges are younger than the circumferential ridges, indicating a change in the regional stress field over time. Hesperia Planum also reveals ridge rings (**Figure 6**) that are interpreted to form over buried impact craters (e.g., Watters, 1993). Hesperia Planum deposits are <3 km thick (Ivanov et al., 2005) so it is unlikely these wrinkle ridges deform the entire lithosphere.

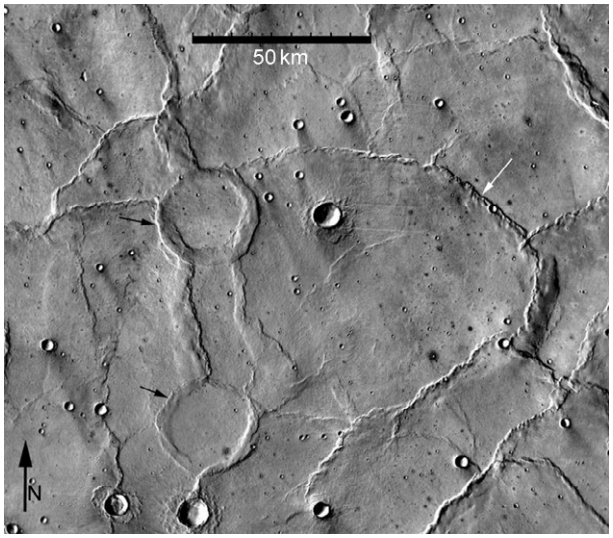


Figure 6 Small, circular ridge rings (black arrows) and a large, ellipsoid ridge ring (white arrow) in Hesperia Planum, Mars. Note intersecting wrinkle ridges throughout the image. Image centered at 115.3° E, 19.0° S. THEMIS daytime IR mosaic rendered using JMars. Image courtesy of NASA/JPL/ASU.

Similar to the Moon, Mars also displays lobate scarps that are interpreted to be lithospheric scale thrust faults (Schultz and Tanaka, 1994). They are larger than wrinkle ridges (both longer and taller), and shortening is estimated to be one to two orders of magnitude greater along lobate scarps than along wrinkle ridges (Watters, 1993).

Graben on Mars range from a few meters across to kilometers across. The interpretation of high-resolution (<3 m per pixel) images shows evidence for erosion and deposition within narrow graben, suggesting that the primary morphology could be obscured (Golombek and Phillips, 2010). The subsurface character of these narrow graben remains unclear. Bounding faults that dip inward at ~60° would intersect at some depth beneath the surface that may represent a mechanical discontinuity. Alternatively, one fault could terminate against a deeper master fault (Schultz et al., 2007). It has been proposed that most martian graben are the expression of near-surface dikes (e.g., Ernst et al., 2001; Mege and Masson, 1996; Scott and Wilson, 2002; Wilson and Head, 2002). In contrast, pit crater chains on Mars – which appear to be genetically related to some narrow graben (Wyrick et al., 2004) – are formed primarily by dilational normal faulting.

10.09.4.2 Mars' Volcanism

To date, all available data indicate that volcanism on Mars is mafic and probably basaltic. Evidence from remote sensing data, volcanic morphology, Mars' meteorites, and in situ analyses of Martian rocks all support basaltic volcanism (**Table 3**). Remote sensing detection of possible evolved (silica-rich) lavas in Nili Patera is best explained by being hydrothermal deposits rather than lavas or igneous intrusions (Skok et al., 2010). Mars has a thin (<600 Pa at mean planetary radius, or mpr) atmosphere (roughly 0.6% of Earth's) that is 95% CO_2 (NASA, 2010). Although liquid water is not currently stable on the surface of Mars, there is abundant morphological (e.g., Carr, 1996) and geochemical (e.g., Morris et al., 2010) evidence that liquid water existed on the surface in the past. Water ice is found on Mars today, being most abundant in the polar regions (Bibring et al., 2004; Clark and McCord, 1982) and in the subsurface (Schorghofer and Forget, 2012). Thus, when investigating volcanism on Mars, it is important to look for the evidence of interaction between magma (lava) and water (ice).

There are three main volcanic provinces on Mars: (1) the Tharsis region, which boasts both the tallest (Olympus Mons) and widest (Alba Mons) volcanoes in the solar system; (2) the Elysium region; and (3) the circum-Hellas region, which contains some of the oldest volcanoes on Mars. In addition, Mars appears abundantly covered in 'ridged plains,' which are interpreted to be thinly layered basalt flows (Greeley and Guest, 1987; Scott and Tanaka, 1986) that were largely emplaced around the same time (~3.2 Ga). Effusive eruptions appear to have been more common than explosive eruptions, although it is possible that the explosive eruptions have not been fully identified yet (e.g., Kerber et al., 2011).

Both the Tharsis and Elysium regions are dominated by shield volcanoes. The Tharsis region contains four large shield volcanoes: Olympus, Ascraeus, Arsia, and Pavonis Montes, as well as a low shield volcano (Albus Mons) and numerous smaller shield volcanoes (**Figure 5**). 'Shield' volcanoes are so

named because they are characterized by shallowly sloping flanks (4°–11°) that, in profile, look like a Viking shield resting on its handle. The shields within the Tharsis region are morphologically similar to terrestrial shield volcanoes (such as Mauna Loa volcano, Hawaii), complete with summit caldera complexes and volcanic rift zones, but are orders of magnitude larger than terrestrial volcanoes. This size difference is likely because Mars does not have plate tectonics, and so, the lithosphere is never moved away from the underlying melt source. The Elysium region contains Elysium Mons and a number of smaller shields. By analogy with terrestrial basaltic shield volcanoes, it is assumed that these martian shields were fed by an underlying mantle plume (on Earth, a 'hot spot').

In contrast, the volcanoes of the circum-Hellas region have flank slopes even shallower than shield volcanoes (<2°), and their flanks are dissected by radial channels. (Note that prior to 2010, these volcanoes are referred to in the literature as 'paterae'; after 2010, 'patera' refers to the summit caldera complexes of these low-lying volcanoes and the volcanic constructs are called 'montes' (IUGG).) As a group, the volcanoes around Hellas are also called 'highland patera' (Greeley and Spudis, 1981) because they are located in the southern highlands. The low flank slopes and the easily eroded nature of the shield materials have led to the suggestion that these volcanoes experienced explosive eruptions and that the shields are composed primarily of pyroclastic materials (e.g., Crown and Greeley, 1993; Greeley and Crown, 1990). Modeling results suggest that the observed distribution of shield materials could be generated by either explosive decompression of magmatic gases or magma–water–ice interactions as rising magma encountered ground ice or groundwater (Greeley and Crown, 1990; Gregg and Farley, 2006). Impact crater size–frequency distributions and stratigraphy indicate that the explosive activity is ancient (>3.5 Ga) (Williams et al., 2007, 2008). Younger lava flows (as young as ~1.5 Ga) were subsequently erupted from both Hadriacus and Tyrrhenus Montes (Werner, 2009; Williams et al., 2007, 2008). Thus, these volcanoes experienced a change in their volcanic styles with time. Unlike the Tharsis and Elysium regions, the circum-Hellas volcanics may be directly related to the formation of the Hellas basin (e.g., Crown and Greeley, 1993; Greeley and Crown, 1990): it is possible that deep fractures generated by the impact event, combined with mantle upwelling in response to lithospheric thinning, may have created preferred pathways for magma to reach the surface.

Volcanic plains are widespread on Mars (Greeley and Guest, 1987; Scott and Tanaka, 1986) and are interpreted to be composed of low-viscosity basalts (cf. Greeley et al., 2005). Head et al. (2002) proposed that in addition to the visible plains material, there are volcanic plains underlying most (if not all) of the northern lowlands. The type example of these volcanic plains on Mars is Hesperia Planum (**Figure 6**), characterized by a flat surface lacking obvious vents and lava tubes/channels and deformed by wrinkle ridges. Morphologically, volcanic plains are most similar to terrestrial flood basalts. Globally, these volcanic plains are ancient, emplaced prior to about 3.5 Ga (Werner, 2009).

Impact crater size–frequency distributions strongly support that most volcanism on Mars is ancient: the major volcanoes were close to their present size by 3.6 Ga (Werner, 2009). Determining whether volcanism on Mars remains 'active' depends entirely on the definition of 'active.' Terrestrial volcanoes are considered to be active if they have erupted in the past 10 000 years, based on the easily observed evidence left by retreating glaciers at that time. There is no similar standard for Mars. However, lava flows that are as young as 1 Ma have been identified (Hauber et al., 2011) through mapping and impact crater size–frequency distributions.

Evidence for lava–water interactions on Mars remains sparse and elusive, although the planetary conditions would favor these interactions in the past (Head and Wilson, 2002). Lava flows with thermal fracture patterns called 'columnar jointing' have been observed in impact crater walls (Milazzo et al., 2009); on Earth, the formation of columnar jointing is enhanced by the presence of meteoric water (Grossenbacher and McDuffie, 1995). Rootless cones on Earth form when basaltic lava flows over water-saturated ground, and these have been suggested to exist on specific martian lava flows (Fagents et al., 2002). Flank slopes of small (<50 km basal diameter) martian shields increase poleward, and Garvin et al. (2000) suggested that magma interaction with groundwater or ground ice was responsible for the poleward morphological shift. Allen (1979) and Hodges and Moore (1994) used photointerpretation of Viking Orbiter images to identify constructs on Mars that are similar to those formed by subglacial basaltic eruptions on Earth. Gregg et al. (2007) suggested that fluid lavas within Gusev crater (Greeley et al., 2005) ponded against an ice-rich deposit that has subsequently been removed. Chapman (2002) argued that layered mounds within Vallis Marineris were generated by subglacial eruptions. In spite of evidence for glaciers on the flanks of the Tharsis region volcanoes in the past (e.g., Fastook et al., 2008), unequivocal evidence for lava–ice interactions there remains elusive.

10.09.5 Venus

Venus is almost the same size and bulk density as Earth (**Table 1**). Notably, the surface of Venus has an average temperature of ~450 °C and a pressure of 9.6×10^6 Pa at mean planetary radius; the atmosphere is mostly CO_2. There is no evidence for water vapor in the atmosphere (e.g., Donahue and Russell, 1997) and liquid water is unstable at Venus' surface; it is likely that the Venusian lithosphere and interior are also anhydrous (Kaula, 1990, 1995; Phillips et al., 1997). The dry nature of Venus' geologic materials influences their mechanical and physical properties (e.g., Lenardic et al., 1995; Phillips et al., 1997). It may be the lack of water on Venus that is largely responsible for the differences in tectonic and volcanic behavior on Earth and Venus. The limited information available for the composition of the Venusian surface comes from modeling results, measurements made by the Venera landers, and inferences based on volcanic morphology (e.g., Baker et al., 1992, 1997; Gregg and Greeley, 1993; Nimmo and McKenzie, 1998). Overwhelmingly, the evidence points to basaltic volcanism, but there are local exceptions (discussed in the succeeding text).

To understand the scientific context of Venusian tectonism and volcanism, it is necessary to grasp the following concepts: First, there is a lack of impact craters on the surface of Venus

(only 967 on the entire planet (USGS Astrogeology Branch, 1998)), indicating that the average surface age is ~750 Ma (McKinnon et al., 1997). Second, there remains a controversy about the mechanisms responsible for such a young surface. Venus does not have plate tectonics, and so, that cannot be the resurfacing mechanism (e.g., McGill et al., 2010). Two end-member hypotheses have been proposed to explain Venus' young surface: (1) catastrophic resurfacing and (2) equilibrium resurfacing (e.g., Basilevsky et al., 1997; Phillips et al., 1992; Strom et al., 1994). In catastrophic resurfacing, the entire surface of Venus is reformed in one fell swoop, possibly through global lithospheric overturn. In equilibrium resurfacing, volcanism and tectonism act throughout space and time to pave a little bit of the surface with lava here and deform a bit of the surface through tectonism there; over time, the entire planet is eventually resurfaced. Regardless, Venus has a geologically young surface compared to Mars, the Moon, and Mercury.

This begs the question: Is Venus 'active' volcanologically or tectonically? The data available for the Venusian surface were collected by synthetic aperture radar (SAR) or by Earth-based radar to see through Venus' optically thick cloud cover. The resolution of the Magellan spacecraft's SAR instrument was, at best, 75 m per pixel (Wall et al., 1995; Young, 1990); Earth-based radar is on the order of ~2 km per pixel. The Magellan mission collected data from X to Y. During that time, at the best available resolutions, no observable change was detected on the surface. However, the European Space Agency craft Venus Express is currently orbiting Venus and measuring atmospheric properties. Upon arrival at Venusin 2006, the instruments detected a surge in the amount of atmospheric SO_2, which then decayed over time (Marcq et al., 2013); this is consistent with a volcanic eruption injecting SO_2 into the atmosphere. In the absence of information about changes in surface morphology, however, the hypothesis of active volcanism on Venus cannot be disproven.

10.09.5.1 Venus Tectonism

The following tectonic terranes have been identified on Venus: (1) plains, (2) volcanic rises, (3) crustal plateaus and tesserae, (4) coronae, and (5) chasmata (canyons) (**Figure 7**; Hansen et al., 1997; McGill et al., 2010). In the absence of plate tectonics, most of these features are explained by either mantle upwelling, downwelling, or some combination of the two.

Plains are stratigraphically superposed on other tectonic terranes and so are locally the youngest materials. They are interpreted to be volcanic plains, composed of thinly layered lava flows; this interpretation is supported by images of the Venusian surface taken by the Venera landers (e.g., Garvin et al., 1984; Surkov et al., 1983). Plains are typically deformed by wrinkle ridges, graben, lineaments, and/or polygonal lineaments. Although lineaments are not tectonic landforms per se, they are locally, at least, likely to be expressions of a small (subresolution) structural feature such as a fracture or fault. Wrinkle ridges on Venus, imaged by radar, look different from those on the other planets: Magellan SAR cannot resolve a broad arch and a superposed sinuous ridge. Instead, wrinkle ridges are expressed by a sinuous bright line (**Figure 8**; note that Magellan images are processed so that bright = high radar backscatter) and are interpreted to represent compressional forces. Commonly, parallel wrinkle ridge sets extend over hundreds to thousands of kilometers, implying a similar stress field over those distances.

Only graben at least 225 m wide can be resolved on the highest-resolution Magellan SAR data; it is inferred that where graben are found parallel with lineaments, the lineaments are also graben but are too small to be completely resolved. Graben and lineaments in radiating swarms, or nova, are abundant on Venus (Head et al., 1992). Based on their morphological similarities to radiating dike swarms on Earth, the Venusian novas are interpreted to be the surface manifestation of radiating dike swarms that did not quite reach the

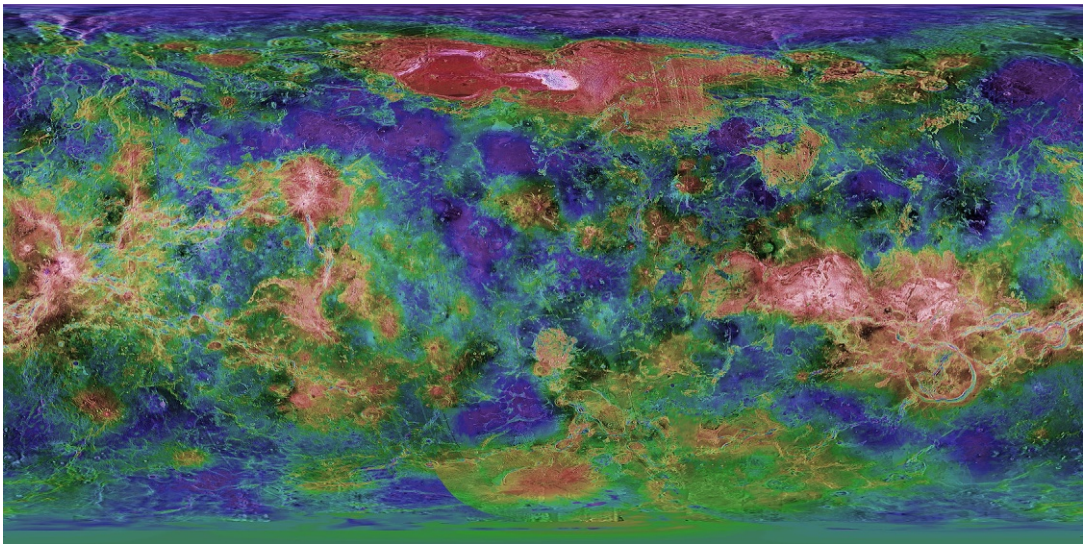


Figure 7 Global Venus topography and SAR. All Magellan images are processed so that bright = high radar backscatter; dark = low radar backscatter. Pinks and whites are topographic highs; blues are lows. (see http://astrogeology.usgs.gov/search/details/Venus/Magellan/RadarProperties/Colorized/Venus_Magellan_C3-MDIR_ClrTopo_Global_Mosaic_6600m/cub).

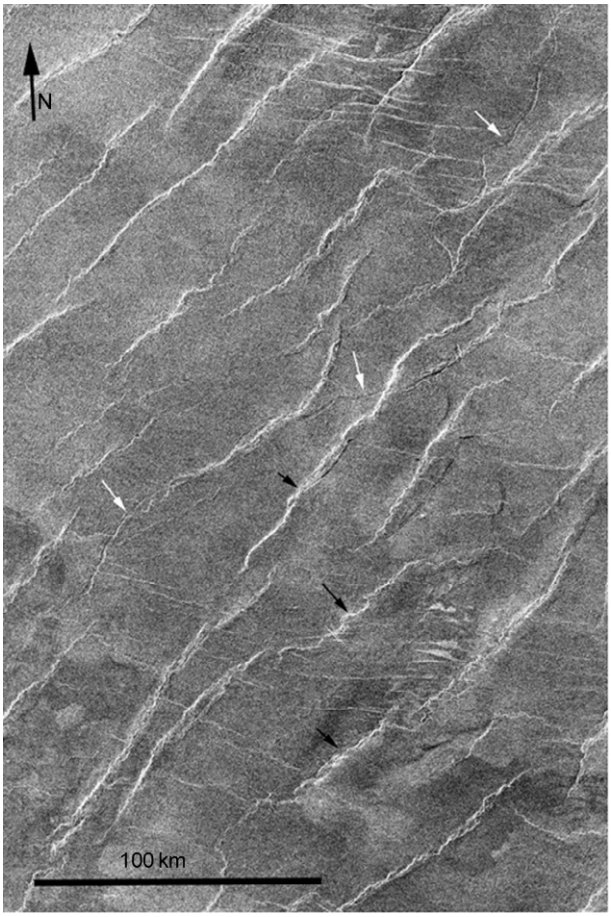


Figure 8 Wrinkle ridges (black arrows) on Venusian plains; image centered at 172° E, 34° S. A sinuous volcanic channel is marked by the white arrows. Image courtesy of NASA/JPL; rendered using mapaplanet.com.

surface to erupt (Grosfils and Head, 1994). Similarly, most of the graben on Venus are interpreted to be the expression of near-surface volcanic dikes (McGill et al., 2010).

Locally, plains are deformed by polygons. These may be constructed of wrinkle ridges (compression) or graben (extension) or of fractures; individual lineaments are commonly too narrow to resolve completely. The precise origin of these polygons remains unclear.

Ridge belts (formally known as 'dorsa') are concentrated zones of compressional deformation within plains materials. They are characterized by a broad (10^2 km) topographic arch, superposed with wrinkle ridges and locally folds (e.g., Banerdt et al., 1997; Zuber, 1990). With one exception (McGill et al., 2010), ridge belts are older than the surrounding plains. Fracture belts are also characterized by a broad topographic arch containing fractures and graben and so are concentrated zones of extensional deformation. They appear to be composed of deformed plains materials.

Ten volcanic rises have been identified to date on Venus, which are 1400–2500 km across, are topographically high (0.5–2.5 km tall), and are spatially associated with some of the largest volcanoes on Venus. Volcanic rises are interpreted to form over mantle plumes, based on topography and gravity data (Phillips and Malin, 1983; Senske et al., 1992); the large compensation depths of some volcanic rises may indicate they are still being actively supported (Smrekar, 1994).

Tessera and crustal plateaus are both topographically higher than the surrounding materials and are older than the surrounding plains. Crustal plateaus are roughly circular regions 1000–3000 km across that are 0.5–4 km above the adjacent plains, and most contain tessera terrain (Hansen et al., 1999; McGill et al., 2010). Compared with volcanic rises, they have shallow apparent compensation depths, suggesting isostatic compensation (McGill et al., 2010). They formed via mantle upwelling (Grimm and Phillips, 1991; Phillips et al., 1991) or downwelling (Bindschadler and Parmentier, 1990; Lenardic et al., 1995). Ishtar Terra is a unique expression of a crustal plateau on Venus (Hansen et al., 1997): it is a large central plain (containing volcanic calderas) surrounded by mountain belts and Maxwell Montes, the tallest mountain on Venus (Bindschadler et al., 1992; Hansen et al., 1997; Keep and Hansen, 1994; Roberts and Head, 1990; Smrekar and Solomon, 1992).

According to the Gazetteer of Planetary Nomenclature (http://planetarynames.wr.usgs.gov/DescriptorTerms), tesserae are 'tile-like, polygonal terrain' but this description does not fully encompass the complicated nature of Venusian tesserae. Tesserae are characterized by multiple intersecting sets of linear to arcuate tectonic features and a high surface roughness relative to the surrounding plains materials (Hansen et al., 1999). Locally, tesserae terrains are the stratigraphically lowest materials, but that does not require that globally all tesserae are the same age (cf. McGill et al., 2010). Given the range of tectonic features and their orientations found in deforming tesserae, it is likely that different patches of tessera represent different episodes of deformation and possibly distinct starting materials (Hansen and Willis, 1996).

Coronae are circular to ovoid features (Basilevsky et al., 1986) 50–260 km across (Glaze et al., 2002) that are typically surrounded by concentric ridges and fractures. Most coronae display volcanic processes: they may contain volcanoes and/or lava flows. They are interpreted to be caused by buoyant mantle material (Hansen, 2003; Smrekar and Stofan, 1997; Stofan et al., 1991) rising from shallower depths than those plumes associated with volcanic rises. Interestingly, coronae are not randomly distributed across the planet, but are concentrated in the Beta–Atla–Themis region (**Figure 9**; Squyres et al., 1993) and tend to be spatially associated with canyon systems (chasmata).

Chasmata are canyons, interpreted to be tectonic rifts. They are <7 km deep and 10^2–10^3 km long (Solomon et al., 1992). The chasmata contain multiple lineaments that can crosscut and extend beyond the rift; although most lineaments are too small to resolve, locally one can be identified as a trough and so all these lineaments are interpreted to be graben or normal faults (McGill et al., 2010). Individual features, such as impact craters (**Figure 10**; Solomon et al., 1992) and volcanoes (Stofan et al., 2005), have been rifted apart by chasmata, allowing for extension to be measured: 10–20 km locally (Connors and Suppe, 2001; Rathbun et al., 1999; Solomon et al., 1992).

10.09.5.2 Volcanism on Venus

The majority of central-vent volcanoes can be identified as shield volcanoes. Crumpler et al. (1997) had classified shield

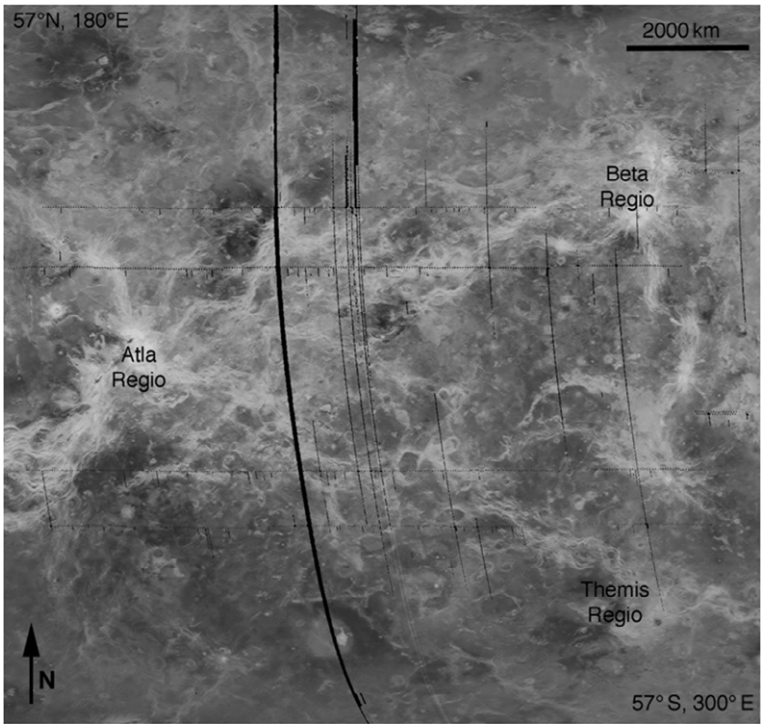


Figure 9 The Beta–Atla–Themis region on Venus, where most of the planet's coronae are located. Image courtesy of NASA/JPL; rendered using mapaplanet.com.

Figure 10 Balch crater, centered at 30° N, 283° W (in Devana Chasma, Beta Regio). Note that the crater has been rifted. Image courtesy of NASA/JPL.

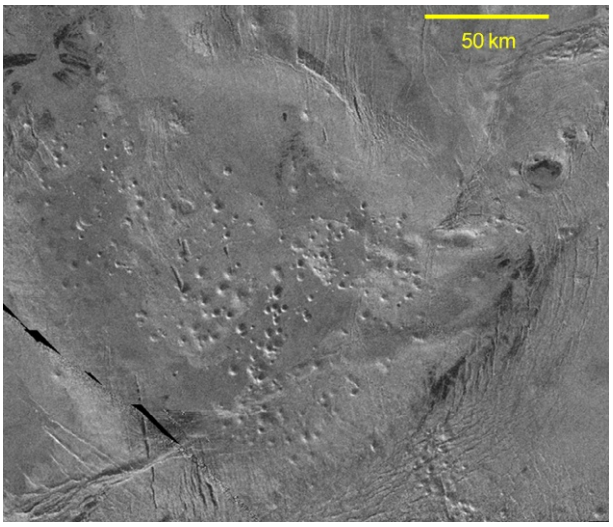


Figure 11 A typical shield field on Venus. Image centered at 78.4° S, 43° E; north is at the top. There are thousands of these fields on Venus. Image courtesy of NASA/JPL/LPI (http://www.lpi.usra.edu/publications/slidesets/venus/slide_26.html).

volcanoes as large (>100 km basal diameter), intermediate, or small (<20 km basal diameter). In addition, there are volcanic features on Venus that are unique among the terrestrial planets.

The most abundant volcanic feature on Venus is small (<25 km basal diameter) shields (Figure 11). There are 10^5–10^6 small shields on Venus, and they typically occur in clusters or groups. Individual shields may contain summit pits. Aubele (2009) had identified two different types of shield clusters: shield fields and shield plains. Shield fields are typically 10^2 km across and contain 10^2 shields, whereas shield plains are 10^3 km across and contain 10^3 shields (Aubele, 2009; Miller and Gregg, 2012). Lava flows from shield fields

commonly bury surrounding wrinkle ridges, but locally lineaments crosscut shields.

Canali are long ($>10^2$ km) sinuous channels found on the Venusian plains and tend to have neither obvious source regions nor sinks. Balta Vallis is the largest example on Venus and is >7000 km long. Along its length, Balta Vallis maintains a consistent width of 3–5 km. The examination of the topography along its length reveals undulations (e.g., Baker et al., 1997) that must have deformed the channel after its emplacement. Long lava flows (10^3 km long) have been identified on the flanks of large shield volcanoes. For comparison, the longest unconfined lava flow yet identified on Earth is the 160 km long Undara flow in Queensland, Australia (see Stephenson et al. (1998) and references therein). (The Grande Ronde flow of the Columbia River Basalts traveled >1000 km from its source, but the lava was confined in a steep and narrow canyon (Mangan et al., 1986).) Many scientists have questioned whether typical tholeiitic basalts could flow such great distances without solidifying, even on Venus, and have therefore considered the possibility of more exotic lava compositions with low melting temperatures and low viscosities (such as carbonatite and sulfur) (e.g., Gregg and Greeley, 1993; Kargel et al., 1994). The main difficulty with the exotic lava hypothesis is that it is untestable: models of flow behavior confirm that carbonatite or sulfur flows could flow hundreds to thousands of kilometers without solidifying, but petrologic models cannot generate the 10^5 km^3 of exotic lavas required to fill the observed channels on Venus. Without more constraints on the composition of Venus, the likelihood of large volumes of exotic lavas cannot be assessed.

Valley networks observed on Venus must be the result of volcanic processes because liquid water cannot exist at the Venusian surface. Baker et al. (1992, 1997) and Komatsu et al. (2001) classified valley networks based on morphology as simple (e.g., canali), complex, compound, or integrated. Valley networks are found in plains materials and may be the source of lava comprising the plains. Some simple valley networks are morphologically similar to lunar sinuous rilles and probably have similar origins. The complex and compound valley networks, consisting of many tributaries, distributaries, and anastamosing reaches, can only be formed by a long-lived (months to centuries) eruption (Gregg and Greeley, 1993). Integrated networks are morphologically similar to valley systems generated by groundwater sapping processes on Earth; they appear to have been formed by collapse caused by subsurface removal of material. On Venus, integrated networks may form as lava tube roof collapse or may represent the presence of a low-viscosity, low-melting-temperature lava on Venus, such as carbonatite or sulfur flows (e.g., Baker et al., 1992, 1997; Kargel et al., 1993; Komatsu et al., 2001).

At the other rheological extreme are 'pancake domes' (**Figure 12**), which are steep-sided, flat-topped volcanoes >20 km across and <5 km tall. Some contain a summit pit. They typically occur in groups and are aligned, suggesting that they were emplaced via a dike. Their large aspect ratio (diameter/height) is similar to the aspect ratio of silica-rich (rhyolite) domes on Earth, such as the Inyo domes in Long Valley, California (Stofan et al., 2000). This observation led to the interpretation that the Venusian domes may also be rhyolitic or at least contain more SiO_2 than typical basalts. The Soviet lander Venera 13 returned measurements of the surface composition consistent with alkali basalts, a type of basalt more evolved than tholeite. Pancake domes are located within the Venera 13 landing elipse (Weitz and Basilevsky, 1993), and it is possible that Venera 13 touched down on top of one of these domes. However, the Venusian domes are 10–100 times more voluminous than terrestrial rhyolite domes, and the petrogenesis of such large quantities of evolved magma in the absence of water or plate tectonics is uncertain. More importantly, detailed analyses of surface roughness and dome topography on Earth and Venus suggest that the Venusian domes are unlike terrestrial silica-rich domes (Stofan et al., 2000).

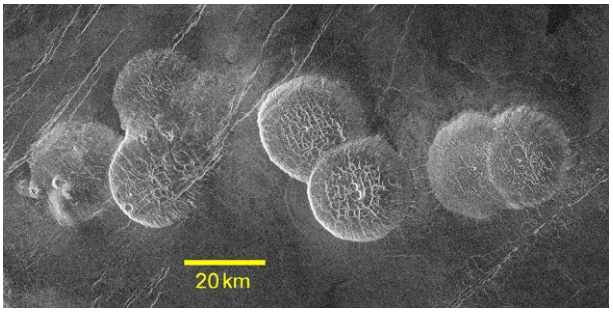


Figure 12 'Pancake' domes on Venus. Image centered at 30° S, 12 °E; north is at the top. Image courtesy of NASA/JPL/LPI (http://www.lpi.usra.edu/publications/slidesets/venus/slide_24.html).

10.09.6 Mercury

Mercury, slightly larger than Earth's Moon (**Table 1**), has the largest iron core relative to planetary radius of all the terrestrial planets. It is also the closest to the Sun, and 'space weathering' (the interaction of solar ions with Mercury's surface) affects the physical properties of surface materials in ways that are not yet fully understood (e.g., Robinson et al., 2008; Zurbuchen et al., 2008). It lacks plate tectonics and is a 'single-plate' planet like the Moon and Mars. In 1974–75, Mariner 10 made 3 flybys of Mercury and succeeded in mapping almost 45% of the planet at a resolution >1 km per pixel. NASA launched the MERcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft in 2004, and as of this writing, it is in orbit around Mercury (see http://www.nasa.gov/mission_pages/messenger/main/index.html), and almost the entire surface has been imaged at resolutions >250 m per pixel.

10.09.6.1 Mercury Tectonics

Tetonic features on Mercury can be broadly categorized as being either 'impact basin-related' or 'distributed' (Watters and Nimmo, 2010; Watters et al., 2009a). The Caloris basin (**Figure 13**) is the largest feature on Mercury and displays one main topographic ring with a diameter of 1550 km. Additional concentric features have been mapped both inside and outside the main ring (Murchie et al., 2008), and these may be additional basin rings. The Caloris basin displays a range of tectonic features that represent 'basin-related' tectonics on Mercury (e.g., Watters et al., 2009b).

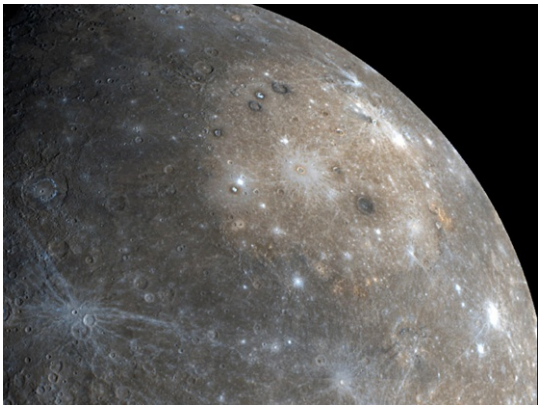


Figure 13 The Caloris basin is the light-colored circular feature near the center of the image and is ~1550 km across. MESSENGER imaged all of Caloris for the first time in 2008. Image courtesy of NASA (http://science.nasa.gov/science-news/science-at-nasa/2008/01oct_mercuryflyby2/).

Tectonic features observed on Mercury include wrinkle ridges, lobate scarps, high-relief ridges, and troughs (interpreted to be graben). Compressional features are far more abundant than extensional ones (Watters and Nimmo, 2010) and are consistent with global contraction (possibly related to the formation of the large iron core). Crosscutting relations suggest that most lobate scarps formed after the emplacement of the smooth plains materials and infilling of Caloris and Rembrandt basins (Solomon et al., 2008; Watters et al., 2005, 2009a) and may be related to local loading (possibly infilling of impact craters with lava).

Lobate scarps are the most common tectonic feature on Mercury (Watters et al., 2009a) and, like those on the Moon and Mars, are interpreted to be thrust faults (**Figure 14**; Nimmo and Watters, 2004; Watters and Nimmo, 2010). They are considered to be 'distributed' and have been observed to crosscut exterior basin fill materials associated with the Caloris basin (cf. Watters et al., 2009b). The oldest materials deformed by lobate scarps are intercrater plains (>4.0 Ga) (Spudis and Guest, 1988; Watters et al., 2009a), and the youngest materials that lobate scarps crosscut are less than ~3.5 Ga old, placing rough constraints on the timing for these features. On Mercury, each lobate scarp represents 1–3 km of shortening (Watters et al., 2009c). The longest lobate scarp yet identified is Enterprise Rupes (centered at 36.5° S, 283.5° W), which is at least 1000 km long (Watters et al., 2009c). Topographic relief across lobate scarps ranges from 0.9 to 2.0 km. Based on the measurements of fault scarp heights and the modeling of thrust fault behaviors, the estimate for Mercury's T_E at time of faulting was on the order of 35–40 km (Watters et al., 2002); recent measurements using the MESSENGER Mercury Laser Altimeter (Cavenaugh et al., 2007) allowed Zuber et al. (2010) to estimate values as high as 60 km for Mercury's T_E.

Wrinkle ridges on Mercury are morphologically similar to those observed on Mars (e.g., Head et al., 2008; Solomon et al., 2008; Strom et al., 1975) (**Figure 15**); even ridge rings (which form over buried impact crater rims) have been identified (Watters and Nimmo, 2010; Watters et al., 2009a). Wrinkle ridges tend to be found in topographic lows and may therefore be related to basin subsidence (Watters 1988, 1993). Importantly, every type of smooth plains yet identified on Mercury (both those interpreted to be volcanic and those interpreted to be impact-related) is deformed by wrinkle ridges (Watters et al., 2009a).

Unlike wrinkle ridges, high-relief ridges are symmetrical in profile and have up to 2.0 km of topographic relief; they can locally transition into lobate scarps. They were first observed in Mariner 10 images and are so-called because they were

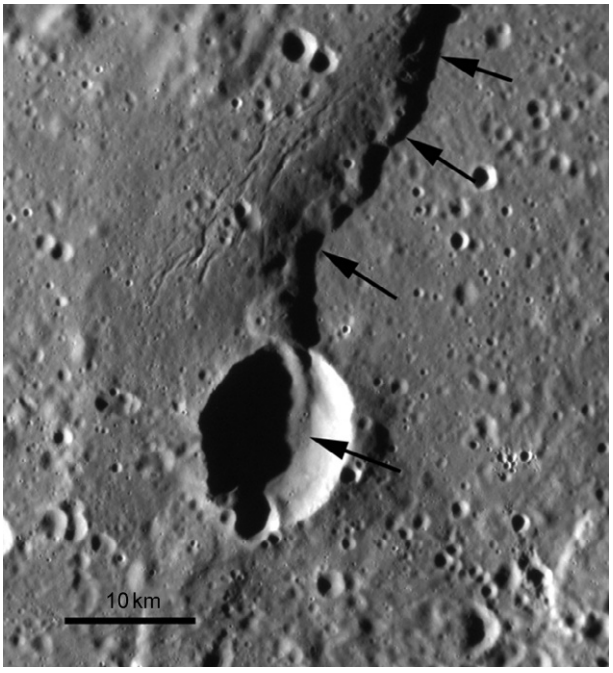


Figure 14 Lobate scarp (black arrows) on Mercury; note that the lobate scarp crosscuts the 12 km diameter impact crater near the center of the image. Image centered at 7.8° N, 151.1° E. MESSENGER narrow-angle camera image courtesy of NASA/JHU/JPL (see http://messenger.jhuapl.edu/gallery/sciencePhotos/pics/Wednesday_September_11.jpg).

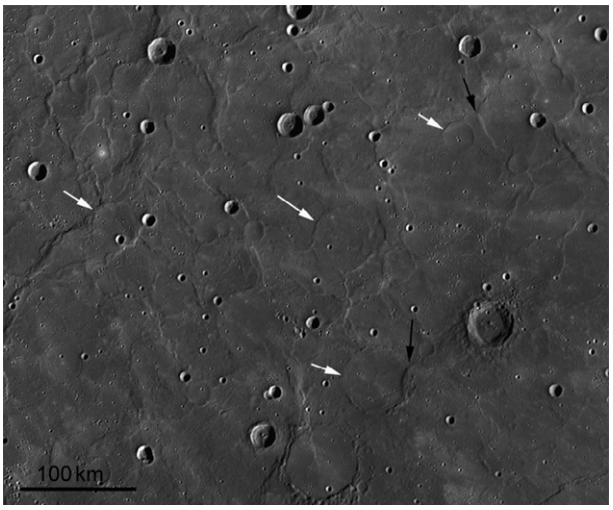


Figure 15 Ridge rings (white arrows) and intersections of wrinkle ridges (black arrows) on Mercury. Image centered at 55.3° N, 32.5 °E. Mercury Dual Imaging System image courtesy of NASA/JHU/APL.

described as ridges with 'significant relief' in the intercrater plains (Dzurisin, 1978). Like wrinkle ridges, high-relief ridges are found within intercrater plains and appear to have formed at approximately the same time as wrinkle ridges (Watters and Nimmo, 2010). Although these features are high-relief, they are only visible in low-sun-angle images or in Mercury Laser Altimeter data (Solomon et al., 2008; Watters et al., 2009a). These observations suggest that high-relief ridges are high-angle reverse faults.

Extensional features are rare on Mercury and were only identified in the eastern half of the Caloris basin from Mariner 10 data (Strom et al., 1975). As of this writing, MESSENGER images reveal linear to arcuate troughs, interpreted as graben, associated with two of the largest impact basins (Caloris and Rembrandt) and three intermediate-sized basins (Raditladi, Rachmaninoff, and Mozart) (Byrne et al., 2013). Caloris contains radial and concentric graben (Murchie et al., 2008; Watters et al., 2009b); Raditladi basin (255 km diameter; 27° N, 119° E) contains concentric graben; and Rembrandt basin (700 km diameter; 33° S, 88° E) contains radially oriented fractures and troughs.

Caloris basin contains Pantheon Fossae (**Figure 16**), a swarm of radiating troughs or graben (Murchie et al., 2008; Solomon et al., 2008; Watters et al., 2009b). On Earth, radiating graben are formed when a mantle plume or diapir impinges on the lithosphere, causing uplift and intrusion of radial dikes (e.g., Ernst et al., 1995); graben form above the rising dikes. By analogy, Pantheon Fossae and the extensional features observed in Rembrandt and Raditladi basins are interpreted to be the result of magmatic uplift sometime after basin formation (similar to floor-fractured craters on the Moon; Wichman and Schultz, 1995). Alternatively, the emplacement of volcanic plains materials outside of the basin may have locally loaded the lithosphere, causing flextural uplift – and resulting extension – in the basin interior (Melosh and McKinnon, 1988). Watters et al. (2005, 2009b) supported lateral flow of the lower crust toward the basin interior, caused by differences in elevation and crustal thickness.

Interestingly, morphologically similar trough (interpreted to be graben) swarms have been identified in Mercurian volcanic plains (Watters et al., 2012b) and are found within wrinkle ridge rings. These swarms, combined with ridge rings, are interpreted to mark the locations of buried impact basins. The juxtaposition of graben encircled by compressional wrinkle ridges likely resulted from a combination of extensional stresses (cooling and thermal contraction of the plains materials) and compressional stresses (related to cooling and contraction of Mercury's interior (Watters et al., 2012b).

10.09.6.2 Mercury Volcanism

Unlike the other terrestrial planets, volcanic edifices are rare on Mercury. Head et al. (2008) characterized a possible volcanic shield volcano within Caloris basin (see also Murchie et al., 2008); Prockter et al. (2010) identified a likely volcanic vent within Rachmaninoff basin (290 km diameter; 28° N, 58° E). Hurwitz et al. (2013) identified possible lava channels on Mercury, although it is not clear if they are constructional or erosional features. The dominant expression of volcanism on Mercury, however, appears to be volcanic plains (Denevi et al., 2013; Head et al., 2008).

Head et al. (2008) examined data collected on Caloris basin from the first MESSENGER flyby and identified volcanic vents and deposits along the southern margin of the interior of the Caloris basin (**Figure 17**). For example, they pointed to a 'kidney-shaped' depression about 20 km long surrounded by relatively a relatively high-albedo deposit covering a roughly

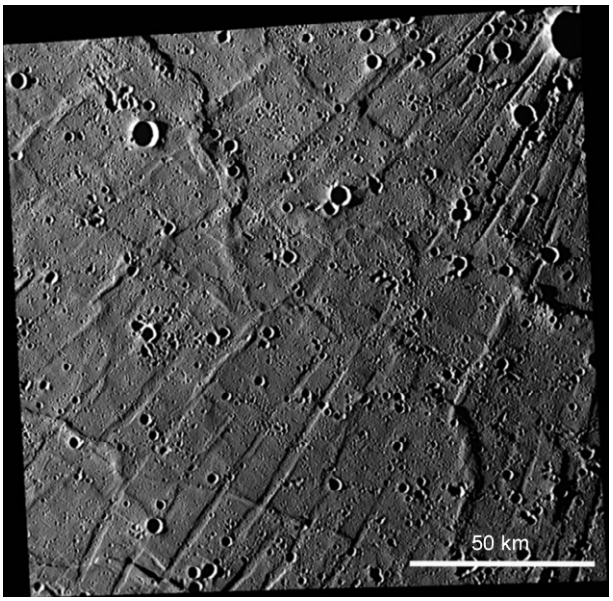


Figure 16 A portion of Pantheon Fossae, one of the few extensional networks on Mercury. Image centered at 25.2° N, 158.9° E. Image acquired by the Mercury Dual Imaging System wide-angle camera; image ID 2213443. Image courtesy of NASA/JHUAPL (see http://messenger.jhuapl.edu/gallery/sciencePhotos/pics/EW0250969390G.map.web.png).

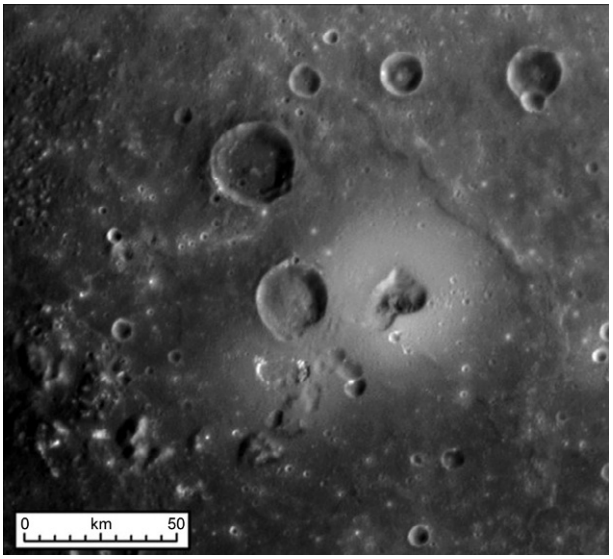


Figure 17 Likely volcanic vents (irregularly shaped depressions in the left of the image) within the rim of Caloris basin. Image centered at 22.0° N, 146.4° E. Image courtesy of NASA/JHUAPL (see http://science1.nasa.gov/media/medialibrary/2008/07/03/03jul_mercuryupdate_resources/kidney.jpg).

circular area up to ~50 km from the depression. They interpreted this to be a volcanic construct (Prockter et al., 2010). Smooth plains materials in and around Caloris basin are also interpreted to be volcanic (Murchie et al., 2008); recent studies (Ernst et al., 2013; Klimczak et al., 2013) suggest that the volcanic fill within the Caloris basin is about 2.5–4.0 km thick.

There is a strong morphological similarity between smooth plains deformed by wrinkle ridges on Mercury and on Mars: both planets display intersecting wrinkle ridges systems and ridge rings (**Figure 15**; Head et al., 2008). The morphological similarities between lunar mare-type wrinkle ridges (found only in the mare basalts), martian wrinkle ridges, and mercurian wrinkle ridges support the interpretation that many of the plains on Mercury are volcanic (Denevi et al., 2013).

10.09.7 Summary

The Earth is the only planetary body in the solar system that displays 'plate tectonics,' although Mars may have experienced Earth-like seafloor spreading briefly shortly after planetary formation. In the absence of plate tectonics, the other planetary bodies of the inner solar system display faults and fractures formed by impact- or volcanic-induced stresses or by more global stresses associated with core formation and cooling. Mare-type wrinkle ridges, first identified and characterized on the lunar basalts, are near-surface thrust faults and are the most abundant tectonic feature in the inner solar system.

Volcanism in the inner solar system is dominated by basalt. This is the case even on Earth, where the entire ocean floor is composed of basaltic lava flows. Volcanic behaviors differ widely on the terrestrial planets, from enormous shield volcanoes on Mars to dark mantling deposits on the Moon. The longest channel in the solar system is a lava-formed channel on Venus: Baltis Vallis is longer than the Earth's Nile River.

As we encounter these volcanic and tectonic extremes, it is essential to remember that Earth provides us with only one datum as we seek to understand the processes that shape our world. Only by examining and understanding volcanism and tectonism throughout the solar system can we be truly sure we understand how those processes operate at home.

References

Acuna MH, Connerney JEP, Wasilewski P, et al. (2001) Magnetic field of Mars: Summary of results from the aerobraking and mapping orbits. *Journal of Geophysical Research* 106(E10): 23403–23417.

Aharonson O, Zuber MT, and Rothman DH (2001) Statistics of Mars' topography from MOLA: Slopes, correlations and physical models. *Journal of Geophysical Research* 106: 20527–20546.

Allen CC (1979) Volcano–ice interactions on Mars. *Journal of Geophysical Research* 84(B14): 8048–8059.

Andrews-Hanna JC, Zuber MT, and Hauck SA II (2008) Strike-slip faults on Mars: Observations and implications for global tectonics and geodynamics. *Journal of Geophysical Research* 113: E08002. http://dx.doi.org/10.1029/2007JE002980.

Antonenko I, Head JW, Mustard JF, and Hawke BR (1994) Criteria for the detection of lunar cryptomaria. *Earth, Moon, and Planets* 69(2): 141–172.

Aubele JC (2009) Shield fields and shield plains on Venus: Contrasting volcanic units exemplified in Shimti Tessera (V-11) and Vellamo Planitia (V-12) Quadrangles. In: *40th Lunar and Planetary Science Conference*, Abstract #2396.

Baker VC, Komatsu G, Gulick VC, and Parker TJ (1997) Channels and valleys. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 757–797. Tucson, AZ: University of Arizona Press.

Baker VC, Komatsu G, Parker TJ, Gulick VC, Kargel JS, and Lewis JS (1992) Channels and valleys on Venus: Preliminary analysis of Magellan data. *Journal of Geophysical Research* 97(E8): 13421–13444.

Banerdt WB, McGill G, and Zuber MT (1997) Plains tectonics on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 901–930. Tucson, AZ: University of Arizona Press.

Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1047–1084. Tucson, AZ: University of Arizona Press.

Basilevsky AT, Pronin AA, Ronca LB, Krlyuchkov VP, Sukhanov AL, and Markov MS (1986) Styles of tectonic deformation on Venus: Analysis of Venera 15 and 16 data. *Proceedings of the 16th Lunar and Planetary Science Conference*. *Journal of Geophysical Research* 91: D399–D411.

Belleguic V, Lognonne P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005. http://dx.doi.org/10.1029/2005JE002437.

Best MG and Christiansen EH (2001) *Igneous Petrology*. Malden, MA: Blackwell Science, Inc., 458 pp.

Beuthe M, Le Maistre S, Rosenblatt P, Patzold M, and Dehant V (2012) Density and lithospheric thickness of the Tharsis Province from MEX MaRS and MRO gravity data. *Journal of Geophysical Research* 117: E04002. http://dx.doi.org/10.1029/2011JE003976.

Bibring JP, Langevin Y, Poulet F, et al. (2004) Perennial water ice identified in the south polar cap of Mars. *Nature* 428(6983): 627–630.

Binder AB (1982) Post-Imbrian global lunar tectonism: Evidence for an initially totally molten Moon. *Earth, Moon, and Planets* 26: 117–133.

Binder AB and Gunga H-C (1985) Young thrust-fault scarps in the highlands: Evidence for an initially totally molten Moon. *Icarus* 63: 421–441.

Bindschadler DL and Parmentier EM (1990) Mantle flow tectonics: The influence of a ductile lower crust and implications for the formation of topographic uplands on Venus. *Journal of Geophysical Research* 95: 21329–21344.

Bindschadler DL, Schubert GG, and Kaula WM (1992) Coldspots and hotspots: Global tectonics and mantle dynamics of Venus. *Journal of Geophysical Research* 97: 13495–13532.

Budney CJ and Lucey PG (1998) Basalt thickness in Mare Humorum: The crater excavation method. *Journal of Geophysical Research* 103(E7). http://dx.doi.org/10.1029/98JE01602.

Burov EB and Diament M (1995) The effective elastic thickness (Te) of continental lithosphere: What does it really mean? *Journal of Geophysical Research* 100(B3): 3905–3927. http://dx.doi.org/10.1029/94JB02770.

BVSP (Basaltic Volcanism Study Project) (1981) *Basaltic Volcanism on the Terrestrial Planets*. New York: Pergamon Press, 1286 pp.

Byrne PK, Klimczak C, Prockter LM, et al. (2013) The origin of graben and ridges in Rachmaninoff, Raditladi, and Mozart basins, Mercury. *Journal of Geophysical Research: Planets* 118(1): 47–58.

Carr MH (1996) *Water on Mars*. New York: Oxford University Press, 248 pp.

Carr MH and Head JW (2010) Geologic history of Mars. *Earth and Planetary Science Letters* 294: 185–203.

Cavenaugh JF, et al. (2007) The Mercury Laser Altimeter instrument for the MESSENGER mission. *Space Science Reviews* 131: 451–480.

Chapman MG (2002) Layered, massive and thin sediments on Mars: Possible Late Noachian to Early Amazonian tephra? In: Smellie JL and Chapman MG (eds.) *Volcano–Ice Interaction on Earth and Mars*, pp. 273–294. London: Geological Society, Special Publication No. 202.

Citron RI and Zhong S (2012) Constraints on the formation of the Martian crustal dichotomy from remnant crustal magnetism. *Physics of the Earth and Planetary Interiors* 212–213: 55–63.

Clark RN and McCord TB (1982) Mars residual north polar cap: Earth-based spectroscopic confirmation of water ice as a major constituent and evidence for hydrated minerals. *Journal of Geophysical Research* 87(B1): 367–370. http://dx.doi.org/10.1029/JB087iB01p00367.

Connors C and Suppe J (2001) Constraints on magnitude of extension on Venus from slope measurements. *Journal of Geophysical Research* 106: 3237–3260.

Crown DA and Greeley R (1993) Volcanic geology of Hadriaca Patera and the eastern Hellas region of Mars. *Journal of Geophysical Research* 98(E2): 3431–3451.

Crumpler LS, Aubele JC, Senske DA, Keddie ST, Magee KP, and Head JW (1997) Volcanoes and centers of volcanism on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 697–756. Tucson, AZ: University of Arizona Press.

Denevi BW, Ernst CM, Whitten JL, et al. (2013) The volcanic origin of a region of intercrater plains on Mercury. In: *44th Lunar and Planetary Science Conference*, Abstract #1218.

Dombard AJ and Gillis JJ (2001) Testing the viability of topographic relation as a mechanism for the formation of lunar floor-fractured craters. *Journal of Geophysical Research* 106: 27901–27910.

Donahue TM and Russell CT (1997) The Venus atmosphere and ionosphere and their interaction with the solar wind: An overview. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 3–32. Tucson, AZ: University of Arizona Press.

Dzurisin D (1978) The tectonic and volcanic history of Mercury as inferred from studies of scarps, ridges, troughs and other lineaments. *Journal of Geophysical Research* 83: 4883–4906.

Ernst CM, Denevi BW, Murchi SL, et al. (2013) Volcanic plains in Caloris basin: Thickness, timing, and what lies beneath. In: *44th Lunar and Planetary Science Conference*, Abstract #2364.

Ernst RE, Grosfils EB, and Mege D (2001) Giant dike swarms: Earth, Venus and Mars. *Annual Review of Earth and Planetary Sciences* 29: 489–534.

Ernst RE, Head JW, Parfitt E, Grosfils E, and Wilson L (1995) Giant radiating dyke swarms on Earth and Venus. *Earth-Science Reviews* 39(1–2): 1–58.

Fagents SA, Lanagan P, and Greeley R (2002) Rootless cones on Mars: A consequence of lava-ground ice interaction. In: Smellie JL and Chapman MG (eds.) *Volcano–Ice Interaction on Earth and Mars*, pp. 295–318. London: Geological Society, Special Publication No. 202.

Fastook JL, Head JW, Marchant DR, and Forget F (2008) Tropical mountain glaciers on Mars: Altitude-dependence of ice accumulation, accumulation conditions, formation times, glacier dynamics, and implications for planetary spin-axis/orbital history. *Icarus* 198: 305–317.

Freed AM, Melosh HJ, and Solomon SC (2001) Tectonics of mascon loading: Resolution of the strike-slip faulting paradox. *Journal of Geophysical Research* 106: 20603–20620.

Frey HV (2006) Impact constraints on the age and origin of the lowlands of Mars. *Geophysical Research Letters* 33: L08S02. http://dx.doi.org/10.1029/2005GL024484.

Frey HV, Roark JH, Shockey KM, Frey EL, and Sakimoto SEH (2002) Ancient lowlands on Mars. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2001GL013832.

Garvin JB, Head JW, Zuber MT, and Helfenstein P (1984) Venus: The nature of the surface from Venera panoramas. *Journal of Geophysical Research* 89(B5): 3381–3399.

Garvin JB, Sakimoto SEH, Frawley JJ, Schnetzler CC, and Wright HM (2000) Topographic evidence for geologically recent near-polar volcanism on Mars. *Icarus* 145: 648–652.

Glaze LS, Stofan ER, Smrekar SE, and Baloga SM (2002) Insights into corona formation through statistical analyses. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2002JE001904.

Golombek MP (1979) Structure analysis of lunar grabens and the shallow crustal structure of the Moon. *Journal of Geophysical Research* 84(B9): 4657–4666.

Golombek MP, Anderson FS, and Zuber MT (2001) Martian wrinkle ridge topography: Evidence for subsurface faults from MOLA. *Journal of Geophysical Research* 106(E10): 23811–23821.

Golombek MP and Phillips RJ (2010) Mars tectonics. In: Watters TR and Schultz RA (eds.) *Planetary Tectonics*, pp. 183–232. New York: Cambridge University Press.

Goudy CL, Schultz RA, and Gregg TKP (2005) Coulomb stress changes in Hesperia Planum, Mars, reveal regional thrust fault reactivation. *Journal of Geophysical Research* 110: E10005. http://dx.doi.org/10.1029/2004JE002293.

Greeley R (1971) Lava tubes and channels in the lunar Marius Hills. *Earth, Moon, and Planets* 3: 289.

Greeley R and Crown DA (1990) Volcanic geology of Tyrrhena Patera, Mars. *Journal of Geophysical Research* 95(B5): 7133–7149.

Greeley R, Foing BH, McSween HY Jr., et al. (2005) Fluid lava flows in Gusev crater, Mars. *Journal of Geophysical Research* 110(E5). http://dx.doi.org/10.1029/2005JE002401.

Greeley R and Guest JE (1987) Geologic map of the eastern equatorial region of Mars. United States Geological Survey Miscellaneous Investigation I-1802B, 1:15,000,000.

Greeley R and Spudis PD (1981) Volcanism on Mars. *Reviews of Geophysics and Space Physics* 19(1): 13–41.

Gregg TKP, Briner JP, and Paris K (2007) Ice-rich terrain in Gusev crater, Mars? *Icarus* 192: 348–360.

Gregg TKP and Farley MA (2006) Mafic pyroclastic flows at Tyrrhena Patera, Mars: Constraints from observations and models. *Journal of Volcanology and Geothermal Research* 155(1–2): 81–89.

Gregg TKP and Greeley R (1993) Formation of venusian canali: Considerations of lava types and their thermal behaviors. *Journal of Geophysical Research* 98(E6): 10873–10882.

Grimm RE and Phillips RJ (1991) Gravity anomalies, compensation mechanisms, and the geodynamics of western Ishtar Terra, Venus. *Journal of Geophysical Research* 96: 8305–8324.

Grosfils EB and Head JW (1994) The global distribution of giant radiating dike swarms on Venus: Implications for the global stress state. *Geophysical Research Letters* 21(8): 701–704.

Grossenbacher K and McDuffie S (1995) Conductive cooling of lava: Columnar joint diameter and stria width as functions of cooling rate and thermal gradient. *Journal of Volcanology and Geothermal Research* 69: 95–103.

Grott M and Breuer D (2009) Implications of large elastic thicknesses for the composition and current thermal state of Mars. *Icarus* 201: 540–548. http://dx.doi.org/10.1016/j.icarus.2009.01.020.

Hansen VL (2003) Venus diapirs: Thermal or compositional? *Geological Society of America Bulletin* 115: 1040–1052.

Hansen VL, Banks BK, and Ghent RR (1999) Tessera terrain and crustal plateaus, Venus. *Geology* 27: 1071–1074.

Hansen VL and Willis JJ (1996) Structural analysis of a sampling of tesserae: Implications for Venus geodynamics. *Icarus* 123: 296–312.

Hansen VL, Willis JJ, and Banerdt WB (1997) Tectonic overview and synthesis. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 797–844. Tucson, AZ: University of Arizona Press.

Hartmann WK (2005) Martian cratering 8: Isochron refinement and the chronology of Mars. *Icarus* 174: 294–320. http://dx.doi.org/10.1016/j.icarus.2004.11.023.

Hartmann WK and Neukum G (2001) Cratering chronology and the evolution of Mars. *Space Science Reviews* 96: 165–194.

Hauber E, Broz P, Jagert F, Jodlowski P, and Platz T (2011) Very recent and wide-spread basaltic volcanism on Mars. *Geophysical Research Letters* 38: L10201. http://dx.doi.org/10.1029/2011GL047310.

Head JW III (1976) Lunar volcanism in space and time. *Reviews of Geophysics and Space Physics* 14(2): 265–300.

Head JW, Crumpler LS, Aubele J, Guest JE, and Saunders RS (1992) Venus volcanism: Classification of volcanic features and structures, associations, and global distribution from Magellan data. *Journal of Geophysical Research* 97: 13153–13197.

Head JW III, Kreslavsky MA, and Pratt S (2002) Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *Journal of Geophysical Research* 107: E15004. http://dx.doi.org/10.1029/2000JE001445.

Head JW III, Murchie SL, Prockter LM, et al. (2008) Volcanism on Mercury: Evidence from the first MESSENGER flyby. *Science* 321(69). http://dx.doi.org/10.1126/science.1159256.

Head JW III and Wilson L (1993) Lunar graben formation due to near-surface deformation accompanying dike emplacement. *Planetary and Space Science* 41(10): 719–727.

Head JW III and Wilson L (2002) Mars: A review and synthesis of general environments and geological settings of magma-H_2O interactions. In: Smellie JL and Chapman MG (eds.) *Volcano-Ice Interaction on Earth and Mars*, pp. 27–57. London: Geological Society, Special Publication No. 202.

Hiesinger H, Head JW III, Wolf U, Jaumann R, and Neukum G (2002) Lunar mare basalt flow units: Thickness determined from crater size-frequency distributions. *Geophysical Research Letters* 29(8). http://dx.doi.org/10.1029/2002GL014847.

Hiesinger H, Head JW III, Wolf U, Neukum G, et al. (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JE00198.

Hodges CA and Moore H (1994) Atlas of volcanic landforms on Mars. United States Geological Survey Professional Paper #1534, 194 pp.

Horz F (1978) How thick are lunar mare basalts? In: *Proceedings of the 9th Lunar and Planetary Science Conference*, pp. 3311–3331.

Hurwitz DM, Head JW, Byrne PK, et al. (2013) Investigating the origin of candidate lava channels on Mercury with MESSENGER data: Theory and observations. *Journal of Geophysical Research: Planets* 118: 471–486. http://dx.doi.org/10.1029/2012JE004103.

Ivanov MA, Korteniemi J, Kostama V-P, et al. (2005) Major episodes of the hydrologic history in the region of Hesperia Planum, Mars. *Journal of Geophysical Research* 110(E12). http://dx.doi.org/10.1029/2005JE002420.

James BP, Zuber MT, and Phillips RJ (2013) Crustal thickness and support of topography on Venus. *Journal of Geophysical Research: Planets* 118(4): 859–875.

Kargel JS, Kirk RL, Fegley B, and Treiman AH (1994) Carbonate-sulfate volcanism on Venus? *Icarus* 112: 219–252.
Kargel JS, Komatsu G, Baker VR, and Strom RG (1993) The volcanology of the Venera and VEGA landing sites and the geochemistry of Venus. *Icarus* 103: 253–275.
Kaula WM (1990) Venus: A contrast in evolution to Earth. *Science* 247: 1191–1196.
Kaula WM (1995) Venus reconsidered. *Science* 270: 1460–1464.
Keep M and Hansen VL (1994) Structural history of Maxwell Montes, Venus: Implications for Venusian mountain belt formation. *Journal of Geophysical Research* 99: 26015–26028.
Kerber L, Head JW, Madeleine J-B, Forget F, and Wilson L (2011) The dispersal of pyroclasts from Apollinaris Patera, Mars: Implications for the origin of the Medusae Fossae Formation. *Icarus* 216: 212–220.
Kiefer WS and Li Q (2009) Mantle convection controls the observed lateral variations in lithospheric thickness on present-day Mars. *Geophysical Research Letters* 36(18). http://dx.doi.org/10.1029/2009GL039827.
Klimczak C, Ernst CM, Byrne PK, et al. (2013) Insights into the subsurface structure of the Caloris basin, Mercury, from assessments of mechanical layering and changes in long-wavelength topography. *Journal of Geophysical Research: Planets* 118(10): 2030–2044. http://dx.doi.org/10.1002/jgre.20157.
Komatsu G, Gulick VC, and Baker VR (2001) Valley networks on Venus. *Geomorphology* 37(3–4): 225–240.
Lawrence SJ, Stopar JD, Hawke BR, et al. (2013) LRO observations of morphology and surface roughness of volcanic cones and lobate flows in the Marius Hills. *Journal of Geophysical Research: Planets* 118(4): 615–634. http://dx.doi.org/10.1002/jgre.20060.
Lenardic W, Kaula WM, and Bindschadler DL (1995) Some effects of a dry crustal flow law on numerical simulations of coupled crustal deformation and mantle convection on Venus. *Journal of Geophysical Research* 100: 16949–16957.
Lodders K (1998) A survey of shergottite, nakhlite and chassigny meteorites whole-rock compositions. *Meteoritics and Planetary Science* 33: A183–A190.
Lucchitta BK and Watkins JA (1978) Age of graben systems on the Moon. In: *Proceedings of the 9th Lunar and Planetary Science Conference* (*Geochimica et Cosmochimica Acta*), pp. 3459–3472.
Mangan MT, Wright TL, Swanson DA, and Byerly GR (1986) Regional correlation of Grande Ronde basalt flows, Columbia River basalt group, Washington, Oregon, and Idaho. *Geological Society of America Bulletin* 97(11): 1300–1318.
Marcq E, Bertaux J-L, Montmessin F, and Belyaev D (2013) Variations of sulphur dioxide at the cloud top of Venus's dynamic atmosphere. *Nature Geoscience* 6: 25–28.
Maxwell TA, El-Baz F, and Ward SW (1975) Distribution, morphology and origin of ridges and arches in Mare Serenitatis. *Geological Society of America Bulletin* 86: 1273–1278.
McCauley JF, Carr MH, Cutts JA, et al. (1972) Preliminary Mariner 9 report on the geology of Mars. *Icarus* 17: 289–327.
McGill GE, Stofan ER, and Smrekar SE (2010) Venus tectonics. In: Watters TR and Schultz RA (eds.) *Planetary Tectonics*, pp. 81–120. Cambridge: Cambridge University Press.
McGovern PJ, Solomon SC, Smith DE, et al. (2002) Localized gravity/topography admittance and correlation spectra on Mars: Implications for regional and global evolution. *Journal of Geophysical Research* 107(E12): 5136. http://dx.doi.org/10.1029/2002JE001854 (Correction, *Journal of Geophysical Research* 109: E07007. http://dx.doi.org/10.1029/2004JE002286).
McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 969–1014. Tucson, AZ: University of Arizona Press.
Mege D and Masson P (1996) Amounts of crustal stretching in Vallis Marineris, Mars. *Planetary and Space Science* 44: 749–781.
Melosh HJ (2011) *Planetary Surface Processes*. New York: Cambridge University Press, 500 pp.
Melosh HJ and McKinnon WB (1988) The tectonics of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 374–400. Tucson, AZ: University of Arizona Press.
Meyer C (2005) *Lunar Sample Compendium*. Houston, TX: Lyndon B. Johnson Space Center. http://curator.jsc.nasa.gov/lunar/compendium.cfm.
Milazzo MP, Keszthelyi LP, Jaeger WL, et al. (2009) Discovery of columnar jointing on Mars. *Geology* 37(2): 171–174.
Miller DM and Gregg TKP (2012) Geologic characteristics and stratigraphic relationships of shield fields versus shield plains on Venus. In: *43rd Lunar and Planetary Science Conference*, Abstract #2311.
Montesi LGJ and Zuber MT (2003a) Spacing of faults at the scale of the lithosphere and localization instability: 1. Theory. *Journal of Geophysical Research* 108(B2): 2110. http://dx.doi.org/10.1029/2002JB001923.
Montesi LGJ and Zuber MT (2003b) Spacing of faults at the scale of the lithosphere and localization instability: 2. Application to the central Indian Basin. *Journal of Geophysical Research* 108(B2): 2111. http://dx.doi.org/10.1029/2002JB001924.
Morris RV, Ruff SW, Gellert R, et al. (2010) Identification of carbonate-rich outcrops on Mars by the Spirit Rover. *Science* 329(5990): 421–424. http://dx.doi.org/10.1126/science.1189667.
Murase T and McBirney AR (1970) Viscosity of lunar lavas. *Science* 167: 1491–1493.
Murchie SL, Watters TR, Robinson MS, et al. (2008) Geology of the Caloris basin, Mercury: A view from MESSENGER. *Science* 321(5885): 73–76.
Nakamura Y (1980) Shallow moonquakes: How they compare with earthquakes. In: *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 1847–1853.
Nakamura Y (2003) New identification of deep moonquakes in the Apollo lunar seismic data. *Physics of the Earth and Planetary Interiors* 139: 197–205.
Nakamura Y (2005) Farside deep moonquakes and deep interior of the Moon. *Journal of Geophysical Research* 110: E01001. http://dx.doi.org/10.1029/2004JE002332.
NASA (2007) *Lunar Sample Allocation Guidebook*. Houston, TX: Lyndon B. Johnson Space Center. http://curator.jsc.nasa.gov/lunar/sampreq/LunarAllocHandbook.pdf.
NASA (2010) Mars fact sheet. http://nssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html.
Navrotsky A (1995) Thermodynamic properties of minerals. In: Ahrens TJ (ed.) *Mineral Physics and Crystallography: A Handbook of Physical Constants*, pp. 18–28. Washington, DC: American Geophysical Union.
Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E088002. http://dx.doi.org/10.1029/2004JE002262.
Nimmo F, Hart SD, Korycansky DG, and Agnor CB (2008) Implications of an impact origin for the martian hemispheric dichotomy. *Nature* 453: 1220–1223. http://dx.doi.org/10.1038/nature07025.
Nimmo F and McKenzie D (1998) Volcanism and tectonism on Venus. *Annual Reviews of the Earth and Planetary Science Letters* 26: 23–51.
Nimmo F and Watters TR (2004) Depth of faulting on Mercury: Implications for heat flux and crustal effective elastic thickness. *Geophysical Research Letters* 31: L02701. http://dx.doi.org/10.1029/2003GL018847.
Nyquist LE and Shih C-Y (1992) The isotope record of lunar volcanism. *Geochimica et Cosmochimica Acta* 56: 2213–2234.
Phillips RJ, Grimm RE, and Malin MC (1991) Hotspot evolution and the global tectonics of Venus. *Science* 252: 651–658.
Phillips RJ, Johnson CL, Mackwell SL, Morgan P, Sandwell DT, and Zuber MT (1997) Lithospheric mechanics and dynamics of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1163–1203. Tucson, AZ: University of Arizona Press.
Phillips RJ and Malin MC (1983) The interior of Venus and tectonic implications. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 159–214. Tucson, AZ: University of Arizona Press.
Phillips RJ, Raubertas RF, Arvidson RE, et al. (1992) Impact craters and Venus resurfacing history. *Journal of Geophysical Research* 97(E10): 15923–15948.
Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.
Prockter LM, Ernst CM, Denevi BW, et al. (2010) Evidence for young volcanism on Mercury from the third MESSENGER flyby. *Science* 329(668). http://dx.doi.org/10.1126/science.1188186.
Pruis MJ and Tanaka KL (1995) The Martian northern plains did not result from plate tectonics. *26th Lunar and Planetary Science Conference*, Part 3, pp. 1147–1148.
Rathbun JA, Janes DM, and Squyres SW (1999) Formation of Beta Region, Venus: Results from measuring strain. *Journal of Geophysical Research* 104: 1917–1928.
Roberts CM (2013) *On the Origin of Lunar Sinuous Rilles*. MS Thesis, University at Buffalo, Buffalo, NY.
Roberts KM and Head JW (1990) Western Ishtar Terra and Lakshmi Planum, Venus: Models of formation and evolution. *Geophysical Research Letters* 17: 1341–1344.
Robinson MS, Murchie SL, Blewett DT, et al. (2008) Reflectance and color variations on Mercury: Regolith processes and compositional heterogeneity. *Science* 321(5885): 66–69.
Schaber GG (1973) Lava flows in Mare Imbrium: Geologic evaluation from Apollo orbital photography. In: *Proceedings of the 4th Lunar and Planetary Science Conference*, pp. 73–92.
Schaber GG, Boyce JM, and Moore JH (1976) The scarcity of mappable flow lobes on the lunar maria: Unique morphology of the Imbrium flows. In: *Proceedings of the 7th Lunar and Planetary Science Conference*, pp. 2783–2800.
Schmincke H-U (2004) *Volcanism*. Berlin: Springer-Verlag, 324 pp.
Schorghofer N and Forget F (2012) History and anatomy of subsurface ice on Mars. *Icarus* 220: 1112–1120.

Schubert G, Lingenfelter RE, and Peale SJ (1970) The morphology, distribution and origin of lunar sinuous rilles. *Reviews of Geophysics and Space Physics* 8: 199–224.

Schultz PH (1976) *Moon Morphology: Interpretations Based on Lunar Orbiter Photography*. Austin, TX: University of Texas Press.

Schultz RA (1989) Strike-slip faulting of ridged plains near Valles Marineris: Mars. *Nature* 341: 424–426.

Schultz RA (2000) Localization of bedding plane slip and back thrust faults above blind thrust faults: Keys to wrinkle ridge structure. *Journal of Geophysical Research* 105(E5): 12035–12052.

Schultz RA, Moore JM, Grosfils EB, Tanaka KL, and Mege D (2007) The Canyonlands model for planetary grabens: Revised physical basis and implications. In: Chapman M (ed.) *The Geology of Mars: Evidence from Earth-Based Analogs*, pp. 371–399. Cambridge: Cambridge University Press.

Schultz PH and Spudis PD (1983) The beginning and end of lunar volcanism. *Nature* 302: 233–236.

Schultz PH, Staid MI, and Pieters CM (2006) Lunar activity from recent gas release. *Nature* 444. http://dx.doi.org/10.1038/nature05303.

Schultz RA and Tanaka KL (1994) Lithospheric-scale buckling and thrust structures on Mars: The Coprates rise and south Tharsis ridge belt. *Journal of Geophysical Research* 99: 8371–8386.

Schultz RA and Watters TR (2001) Forward mechanical modeling of the Amenthes Rupes thrust fault on Mars. *Geophysical Research Letters* 28: 4659–4662.

Scott DH and Tanaka KL (1986) Geologic map of the western equatorial region of Mars. United States Geological Survey Miscellaneous Investigation Series I-1802A, 1:15,000,000.

Scott ED and Wilson L (2002) Plinian eruptions and passive collapse events as mechanisms of formation for martian pit chain craters. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2000JE001432.

Senske DA, Schaber GG, and Stofan ER (1992) Regional topographic rises on Venus: Geology of western Eistla Region and comparisons to Beta Regio and Atla Regio. *Journal of Geophysical Research* 97: 13395–13420.

Shearer CK, Hess PC, Wieczorek MA, et al. (2006) Thermal and magmatic evolution of the Moon. In: Jolliff BL, Weczorek MA, Shearer CK, and Neal CR (eds.) *New Views of the Moon, Vol. 60: Reviews in Mineralogy and Geochemistry*, pp. 365–517. Washington, DC: Mineralogical Society of America.

Skok JR, Mustard JF, Ehlmann BL, Milliken RE, and Murchie SL (2010) Silica deposits in the Nli Patera caldera on the Syrtis Major volcanic complex on Mars. *Nature Geoscience* 3: 838–841. http://dx.doi.org/10.1038/ngeo990.

Sleep NH (1994) Martian plate tectonics. *Journal of Geophysical Research* 99: 5639–5655.

Smith DE, Zuber MT, Phillips RJ, et al. (2012) Gravity field and internal structure of Mercury from MESSENGER. *Science* 336(214): 214–217. http://dx.doi.org/10.1126/science.1218809.

Smrekar SE (1994) Evidence for active hotspots on Venus from analysis of Magellan gravity data. *Icarus* 112: 2–26.

Smrekar SE and Solomon SC (1992) Gravitational spreading of high terrain in Ishtar Terra, Venus. *Journal of Geophysical Research* 97: 16121–16148.

Smrekar SE and Stofan ER (1997) Coupled upwelling and delamination: A new mechanism for coronae formation and heat loss on Venus. *Science* 277: 1289–1294.

Snyder GA, Borg LE, Nyquist LE, and Taylor LA (2000) Chronology and isotopic constraints on lunar evolution. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 358–387. Tucson, AZ: University of Arizona Press.

Solomon SC and Head JW (1979) Vertical movement in mare basins: Relation to mare emplacement, basin tectonics and lunar thermal history. *Journal of Geophysical Research* 94: 1667–1682.

Solomon SC and Head JW (1980) Lunar mascon basins: Lava filling, tectonics, and evolution of the lithosphere. *Reviews of Geophysics and Space Physics* 18: 107–141.

Solomon SC, McNutt RL Jr. Watters TR, et al. (2008) Return to Mercury: A global perspective on MESSENGER's first Mercury flyby. *Science* 321(5885): 59–62.

Solomon SC, Smrekar SE, Bindschadler DL, et al. (1992) Venus tectonics: An overview of Magellan observations. *Journal of Geophysical Research* 97: 13199–13255.

Spudis PD (2011) Large shield volcanoes on the Moon. In: *42nd Lunar and Planetary Science Conference*, Abstract #1367.

Spudis PD and Guest JE (1988) Stratigraphy and geologic history of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 118–164. Tucson, AZ: University of Arizona Press.

Squyres SW, Janes DM, Schubert G, et al. (1993) The spatial distribution of coronae and related features on Venus. *Geophysical Research Letters* 20: 2965–2968.

Stephenson PJ, Burch-Johnston AT, Stanton D, and Whitehead PW (1998) Three long lava flows in north Queensland. *Journal of Geophysical Research* 103(B11): 27359–27370.

Stofan ER, Anderson SW, Crown DA, and Plaut JJ (2000) Emplacement and composition of steep-sided domes on Venus. *Journal of Geophysical Research* 105(E11): 26757–26771.

Stofan ER, Bindschadler DL, Head JW, and Parmentier E (1991) Corona structures on Venus: Models of origin. *Journal of Geophysical Research: Planets* 96(E4): 20933–20946.

Stofan ER, Brian AW, and Guest JE (2005) Resurfacing styles and rates on Venus: Assessment of 18 Venusian quadrangles. *Icarus* 173: 312–321.

Strom RG, Schaber GG, and Dawson DD (1994) The global resurfacing of Venus. *Journal of Geophysical Research* 99(E5): 10899–10926.

Strom RG, Trask NJ, and Guest JE (1975) Tectonism and volcanism on Mercury. *Journal of Geophysical Research* 80: 2478–2507.

Surkov YA, Moskalyeva LP, Shcheglov OP, et al. (1983) Determination of the elemental composition of rocks on Venus by Venera 13 and Venera 14 (preliminary results). *Journal of Geophysical Research* 88(S02): A481–A493.

USGS Astrogeology Branch (1998) *Venus Magellan Impact Crater Database*. http://astrogeology.usgs.gov/geology/venus-magellan-crater-database (accessed 2 February 2013).

Wall SD, McConnell SL, Leff CE, Austin RS, Beratan KK, and Rokey MJ (1995) *User Guide to the Magellan Synthetic Aperture Radar Images*. Pasadena, CA: Jet Propulsion Laboratory. NASA Reference Publication 1356. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950018567_1995118567.pdf (accessed 2 February 2013).

Watters TR (1988) Wrinkle ridge assemblages on the terrestrial planets. *Journal of Geophysical Research* 93(B9): 10236–10254.

Watters TR (1992) System of tectonic features common to Earth, Mars and Venus. *Geology* 20: 609–612.

Watters TR (1993) Compressional tectonism on Mars. *Journal of Geophysical Research* 98(E9): 17049–17060.

Watters TR (2003a) Thrust faulting along the dichotomy boundary in the eastern hemisphere of Mars. *Journal of Geophysical Research* 108(E6): 5054. http://dx.doi.org/10.1029/2002JE001934.

Watters TR (2003b) Lithospheric flexure and the origin of the dichotomy boundary on Mars. *Geology* 31: 271–274, http://dx.doi.org/10.1130/0091-7613(2003)031<0271:LFATOO>2.0.CO;2.

Watters TR (2004) Elastic dislocation modeling of wrinkle ridges on Mars. *Icarus* 171: 284–294.

Watters TR, Head JW, Solomon SC, et al. (2009a) Evolution of the Rembrandt impact basin on Mercury. *Science* 324. http://dx.doi.org/10.1126/science.1172109.

Watters TR and Johnson CL (2010) Lunar tectonics. In: Watters TR and Schultz RA (eds.) *Planetary Tectonics*. New York: Cambridge University Press, 518 pp.

Watters TR, McGovern PJ, and Irwin RP III (2007) Hemispheres apart: The crustal dichotomy on Mars. *Annual Review of Earth and Planetary Science* 35: 621–652. http://dx.doi.org/10.1146/annurev.earth.35.031306.140220.

Watters TR, Murchie SL, Robinson MS, et al. (2009b) Emplacement and tectonic deformation of smooth plains in the Caloris basin, Mercury. *Earth and Planetary Science Letters* 285: 309–319.

Watters TR and Nimmo F (2010) The tectonics of Mercury. In: Watters TR and Schultz RA (eds.) *Planetary Tectonics*, pp. 15–80. New York: Cambridge University Press.

Watters TR, Nimmo F, and Robinson MS (2005) Extensional troughs in the Caloris basin of Mercury: Evidence of lateral crustal flow. *Geology* 33: 669–672.

Watters TR, Robinson MS, Banks ME, Tran T, and Denevi BW (2012a) Recent extentional tectonics on the Moon revealed by the Lunar Reconnaissance Orbiter Camera. *Nature Geoscience* 5: 181–185. http://dx.doi.org/10.1038/ngeo1387.

Watters TR, Robinson MS, Beyer RA, et al. (2010) Evidence of recent thrust faulting on the Moon revealed by the Lunar Reconnaissance Orbiter Camera. *Science* 329(936). http://dx.doi.org/10.1126/science.1189590.

Watters TR and Schultz RA (2010) *Planetary Tectonics*. New York: Cambridge University Press, 518 pp.

Watters TR, Schultz RA, Robinson MS, and Cook AC (2002) The mechanical and thermal structure of Mercury's early lithosphere. *Geophysical Research Letters* 29(11): 1542. http://dx.doi.org/10.1029/2001GL014308.

Watters TR, Solomon SC, Klimczak C, et al. (2012b) Extension and contraction within volcanically buried impact craters and basins on Mercury. *Geology* 40: 1123–1126. http://dx.doi.org/10.1130/G33725.1.

Watters TR, Solomon SC, Robinson MS, Head JW, Andre SL, Hauck SA, and Murchie SL (2009c) The tectonics of Mercury: The view after MESSENGER'S first flyby. *Earth and Planetary Science Letters* 285: 283–296. http://dx.doi.org/10.1016/j.epsl.2009.01.025.

Weber RC, Lin P-Y, Garnero EJ, Williams Q, and Lognonné P (2011) Seismic detection of the lunar core. *Science* 331: 309–312.
Weitz CM and Basilevsky AT (1993) Magellan observations of the Venera and Vega landing site regions. *Journal of Geophysical Research* 98(E9): 17069–17097.
Weitz CM and Head JW (1999) Spectral properties of the Marius Hills volcanic complex and implications for the formation of lunar domes and cones. *Journal of Geophysical Research* 104(E8): 18933–18956.
Weitz CM, Head JW, and Pieters CM (1998) Lunar regional dark mantle deposits: Geologic, multispectral, and modeling studies. *Journal of Geophysical Research* 103(E10): 22725–22759.
Werner SC (2009) The global martian volcanic evolutionary history. *Icarus* 201: 44–68.
Wichman RW and Schultz PH (1995) Floor-fractured craters in Mare Smythii and west of Oceanus Procellarum: Implications of crater modification by viscous relaxation and igneous intrusion models. *Journal of Geophysical Research* 100(E10): 21201–21218.
Wieczorek MA (2007) Gravity and topography of the terrestrial planets. In: Spohn T (ed.) *Treatise on Geophysics*, vol. 10, pp. 165–206. Amsterdam: Elsevier.
Wieczorek MA, Jolliff BL, Khan A, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364.
Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339(6120): 671–675. http://dx.doi.org/10.1126/science.1231530.
Wilhelms D (1987) *The Geologic History of the Moon*. Professional Paper 1348, 302 pp. Washington, DC: US Geological Survey.
Williams DA, Fagents SA, and Greeley R (2000) A reassessment of the emplacement and erosional potential of turbulent, low-viscosity lavas on the Moon. *Journal of Geophysical Research* 105(E8): 20189–20205.
Williams DA, Greeley R, Werner SC, et al. (2008) Tyrrhena Patera: Geologic history derived from *Mars Express* High Resolution Stereo Camera. *Journal of Geophysical Research* 113: E11005. http://dx.doi.org/10.1029/2008JE003104.
Williams DA, Greeley R, Zuschneid W, et al. (2007) Hadriaca Patera: Insights into its volcanic history from Mars Express High Resolution Stereo Camera. *Journal of Geophysical Research* 112: E10004. http://dx.doi.org/10.1029/2007JE002924.
Wilson L and Head JW (1981) Ascent and eruption of basaltic magma on the Earth and Moon. *Journal of Geophysical Research* 86(B4): 2971–3001.
Wilson L and Head JW III (2002) Tharsis-radial graben systems as the surface manifestation of plume-related dike intrusion complexes: Models and implications. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JE001593.
Wyrick D, Ferrill DA, Morris AP, Colton SL, and Sims DW (2004) Distribution, morphology and origins of Martian pit crater chains. *Journal of Geophysical Research* 109: E06005. http://dx.doi.org/10.1029/2004JE002240.
Yingst RA and Head JW (1997) Volumes of lunar lava ponds in South Pole-Aitken and Orientale basins: Implications for eruption conditions, transport mechanisms, and magma source regions. *Journal of Geophysical Research* 102: 10909–10931.
Young C (1990) *The Magellan Venus Explorer's Guide*. Pasadena, CA: Jet Propulsion Laboratory. http://www2.jpl.nasa.gov/magellan/guide.html.
Zuber MT (1990) Ridge belts: Evidence for regional- and local-scale deformation on the surface of Venus. *Geophysical Research Letters* 17(9): 1369–1372.
Zuber MT (1995) Wrinkle ridges, reverse faulting, and the depth penetration of lithospheric strain in Lunae Planum, Mars. *Icarus* 114: 80–92.
Zuber MT and Aist LL (1990) The shallow structure of the Martian lithosphere in the vicinity of the ridged plains. *Journal of Geophysical Research* 95(B9): 14215–14230. http://dx.doi.org/10.1029/JB095iB09p14215.
Zuber MT, Montesi LGJ, Farmer GT, et al. (2010) Accommodation of lithospheric shortening on Mercury from altimetric profiles of ridges and lobate scarps measured during MESSENGER flybys 1 and 2. *Icarus* 209: 247–255.
Zurbuchen TH, Raines JM, Gloeckler G, et al. (2008) MESSENGER observations of the composition of Mercury's ionized exosphere and plasma environment. *Science* 32(5885): 90–92.

This page intentionally left blank

10.10 Exogenic Dynamics, Cratering, and Surface Ages

SC Werner, University of Oslo, Oslo, Norway
BA Ivanov, Institute for Dynamics of Geospheres, Moscow, Russia

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by B A Ivanov and W K Hartmann, Volume 10, pp. 207–242, © 2007, Elsevier B.V.

10.10.1	**Introduction**	327
10.10.2	**The Evolution of the Solar System**	328
10.10.3	**Impact Craters: Morphology**	330
10.10.4	**Basics of Impact Cratering Processes**	331
10.10.4.1	Principal Stages of Impact Crater Formation	331
10.10.4.2	Impact Crater Scaling Laws	332
10.10.4.2.1	The concept of coupling	332
10.10.4.2.2	Strength and gravity scaling	333
10.10.4.2.3	Scaling of the impact angle	337
10.10.4.2.4	Impact heating, melting, and vaporization	337
10.10.5	**Cratering Statistics**	338
10.10.5.1	SFD of Impact Craters	339
10.10.5.2	Hartmann Production Function	340
10.10.5.3	Neukum Production Function	341
10.10.6	**Impact Rate Estimates**	342
10.10.6.1	Gravitational Mechanisms	343
10.10.6.2	Nongravitational Mechanisms	343
10.10.6.3	Impact Probability	343
10.10.7	**Extraterrestrial Surface Dating**	346
10.10.7.1	Principles of Surface Dating by Crater Counting	346
10.10.7.2	Uncertainties and Issues of the Crater-Based Surface Dating Method	348
10.10.7.3	Chronology Model Transfer from the Moon to Other Planetary Bodies	350
10.10.7.4	The Earth–Moon System	351
10.10.7.5	Mercury	355
10.10.7.6	Venus	355
10.10.7.7	Mars	355
10.10.7.8	Cratering on Asteroids	355
10.10.7.9	Interpretation of the Small-Sized Crater Statistics	356
10.10.7.10	Effects of Target Properties	356
10.10.7.11	Secondary Cratering	356
10.10.7.11.1	SFD for secondary craters	357
10.10.7.12	Effects of Atmosphere and Geologic Processes	358
10.10.8	**Conclusions**	358
Acknowledgments		358
References		358

10.10.1 Introduction

Exogenic dynamics of planets (in contrast to endogenic dynamics) refers to externally generated dynamical processes, and here, we discuss specifically the exchange of material between the different zones of the solar system and the effects when this material interacts with the surface of a solid planetary body. Material can be transported due to the gravitational perturbation of planets acting on the material rotating around the Sun. This material reaches the solid surface of a planet with high velocity, which ranges from a few km s^{-1} up to 100 km s^{-1}, when their orbits cross each other. If the mass difference is large, the gravitational attraction of the larger body assists in modifying the orbit and it can raise the collision probability under a near-encounter (gravitational focusing). The collision with the planetary solid surface (or with a planetary atmosphere) decelerates the falling body (the *projectile*), converting its kinetic energy into potential energy, heat, light radiation, and shock waves and creating cavities – impact craters – at the surface of the struck planets. High-velocity impacts redistribute ejected material; create a temporary *hot spot* in the crust, melt, fracture, demagnetize rocks; displace them; and sometimes eject a small portion of rocks out of the planet's gravity field. The ejecta can reach another planetary body, providing material exchange between planets. Currently, we find Martian (Martian meteorites belong to a group with the historical abbreviation SNC named after shergottite, nakhlite, and chassignite, now recognized as rocks from Mars (Wood and Ashwal, 1981)) and lunar meteorites, as well as asteroidal fragments, such as the HED meteorites, on Earth. (HED, which stands for

howardite, eucrite, and diogenite that are considered to have originated from the asteroid 4 Vesta (McCord et al., 1970).) The impact site can be modified subsequently, for example, by readjustment due to gravity of impact craters, cooling of impact melt body, or general geologic activity. The modification depends on the size of the planet, its mechanical and thermal structure, and the size of the impact event. The most significant difference between impact site modification in the early stages of planetary evolution and today relates to the cooling of the terrestrial planets with time.

The efficiency of the exogenic dynamics and modification rate of planetary surfaces depend on the frequency of impacts of a given energy and the geologic evolution of the impact site. The frequency of impacts is coupled to the power of the small-body sources and the efficiency of small-body delivery to the target planet. Both parameters did evolve, being different in the first 0.5 Ga after the solar system formation (ancient bombardment rate) and today (the modern bombardment rate). The projectile flux evolution is intimately connected to the history of the solar system formation. The current paradigm of the solar system formation starts with the collapse of the protosolar dust–gas nebula and the ignition of the star (Sun) in the center, followed by events that included the accretion of dust, dust/ice, and gas into sizable bodies in a rotating disk, while gas continuously is removed by solar radiation, as discussed in pioneering work by Safronov (1969) and Wetherill (1978, 1980). The dust/ice ratio in growing bodies is probably controlled by the solar insolation, limiting the presence of volatiles close to the heating Sun. This resulted in the formation of protoplanetary bodies with volatile proportion growing with the distance from the Sun, manifested in the formation of more rocky bodies within the modern Jovian orbit with semimajor axis $a = 5.2$ AU (1 AU – astronomical unit – is equal to the Earth–Sun distance of $\sim 1.5 \times 10^6$ km) and protoplanetary bodies with much larger volatile (frozen gases) content beyond that. The feeding zone in the central part of the solar system produced terrestrial planets (Mercury, Venus, Earth, and Mars), their satellites (the Earth's Moon is the biggest), and a swarm of small bodies, asteroids, between the orbits of Mars and Jupiter by mutual collision and accumulation events of preplanetary small bodies (planetesimals). The processes involved in the growth of planetesimals from dust grains are poorly understood (e.g., Chambers, 2010). Later phases of solar system evolution consider the implications of the giant planets' growth on the terrestrial planet formation and the evolution of the projectile source (e.g., reviewed by Morbidelli et al., 2012a). Jupiter and the other giant planets (Saturn, Uranus, and Neptune) accumulated a huge amount of gases before the total dissipation of the gas portion of the disk. A second belt of planetesimals is found beyond the orbit of Neptune, the Kuiper–Edgeworth Belt, which remained due to a combination of too low disk density and too high accretional timescales. The composition of these planetesimals is supposedly consistent with volatile low-temperature condensates formed at large distances from the Sun (reviewed by Brown, 2012).

10.10.2 The Evolution of the Solar System

The view on the evolution of our solar system has become more complex after the discovery of many very different extrasolar planetary systems (e.g., Morbidelli et al., 2012a), and we will try here to evaluate the implications of these more complex evolution schemes on the method of surface-age determination using the impact crater record of solid surface planetary bodies. The early stage of the planetary formation is controlled by celestial mechanics of small bodies revolving around the Sun. The orbits of the small bodies change due to mutual collisions and (possibly) the drag of gaseous disk residuals. Planetary embryos form in a very short time (e.g., Morbidelli et al., 2012a). The growth of a few main planets results in 'excitation' of the orbits of smaller bodies and, possibly, in mixing of individual planetary feeding zones at the end of the main planet formation phase (see **Chapter 10.16**). After the Sun, Jupiter is the most massive and gravitationally important body in the solar system. Due to its significant mass, Jupiter sets up resonance zones that perturb the motion of other planets, small bodies, and their fragments. Beyond Mars, the presence of growing Jupiter exited the orbital motion of local protoplanetary bodies (the origin of natural satellites is discussed in **Chapter 10.17**). Collision velocities increased, fragmentation began to dominate over accretion, and this limited the size of the largest bodies accumulated in this zone, which becomes later the main asteroid belt (MAB). Currently, the largest body in the MAB is Ceres with a diameter of ~1000 km. The formation of larger, for example, Mars-sized, bodies cannot be excluded, and they could have originally been accumulated here or elsewhere among terrestrial planets but removed soon after by the gravitational perturbation of Jupiter and mutual gravitational interaction between the largest protoplanets (Chambers and Wetherill, 2001). The flux of these bodies together with *leftovers* from the feeding zones of the terrestrial planets created the early heavy bombardment of just-formed crusts on terrestrial planets. The timescale of terrestrial planet formation is up to 100 Ma (reviewed by Kleine et al., 2009). The largest MABs (>~100 km in diameter) are most probably leftovers of the primary population, while the smaller ones are probably fragments, formed during the collisional evolution. Some small bodies leave the MAB into planetary-crossing orbits under the action of orbital resonances with Jupiter and Saturn. The motion of planet-crossing bodies also evolves due to weak resonances with Mars and close encounters with the terrestrial planets. The small-body lifetime of a typical planet-crosser is between 1 and 30 Ma (Gladman et al., 2000), and thus, a steady-state flow of bodies supplied from the MAB is required before one of them strikes the surface of one of the terrestrial planets.

The planet accumulation rate is strongly controlled by the period of one revolution about the Sun. At the distance of the Earth, a body makes one full revolution during 1 year. At the distance of Jupiter, it takes 11 years. At the distance of modern Uranus, the revolution period is as large as about 84 years, and for Neptune, about 165 years. Theoretical accumulation rate estimates based on the distance from the Sun result in a mismatch between the formation time of terrestrial planets and too large time for the formation of Uranus in their today's observed position. Therefore, Zharkov and Kozenko (1990) proposed the hypothesis that permits Uranus and Neptune or the formation of their embryos closer to the Sun with following migration to the modern orbits (numerically tested later by, e.g., Masset and Snellgrove, 2001). In addition to an early migration in the gas envelope, the gravitational interaction with massive icy planetesimal disk is considered to be the main engine for

the post-gas giant planet migration (Fernandez and Ip, 1984; Malhotra, 1993; Tsiganis et al., 2005).

The migration of giant planets, as well as their formation (mass concentration appearance), would 'excite' orbits of small bodies, producing an additional *bombardment epoch* for the inner planets and causing the ejection of parts of small bodies far from the Sun (as suggested in the Nice model by a series of publications: Gomes et al. (2005), Tsiganis et al. (2005), and Morbidelli et al. (2005)). Accordingly, 'leftovers' of the giant planet formation shaped the periphery of the solar system: the spherical shell of comets (cometary nuclei, in a strict sense), the Kuiper–Edgeworth Belt beyond the Neptune orbit, and also the Pluto–Charon system. (The Pluto–Charon system, had it been discovered today, would have been cataloged as a member of the Kuiper–Edgeworth Belt (Bertoldi et al., 2006; Sheppard, 2006.) In 2006, Pluto's official status was changed from 'planet' to 'dwarf planet' (Green, 2006; see **Chapter 10.01** for definitions of planets and dwarf planets).) However, the giant planet evolution in our solar system and in extrasolar systems remains unclear (reviewed by Lissauer and Stevenson, 2007) for modern migration models. The giant planets have to form faster than the terrestrial planets (e.g., Morbidelli et al., 2012a). After the possible heavy bombardment epoch, related to the giant planet migration, the small-body population was mixed, and the most icy planetesimals remained beyond the orbit of Neptune. This group of trans-Neptunian objects or further out Kuiper Belt objects (KBO, discussed in **Chapter 10.20**), are occasional returning comet nuclei and KBOs close to the Sun, and create a small-body bombardment population of comets. Comets, spending considerable time close to the Sun, loose most of their volatiles and can become a subpopulation of dormant or extinct comets, which can develop asteroid-like orbits (e.g., Levison et al., 2009). The proportion of active, dormant, or extinct comets in the total bombardment flux on terrestrial planets is a matter of current debates (Fernandez et al., 2002; Stuart and Binzel, 2004), but most of the comets entering the inner solar system are expected to disrupt (Levison et al., 2002). The first version of the *Nice* (Nice after the town in France, where the model was developed) model (Gomes et al. (2005), Tsiganis et al. (2005), and Morbidelli et al. (2005)), with the conceptional idea of giant planet migration, argued for comets from the outer solar system, restricted to the period of the excitation due to the migration as major projectile population for the terrestrial planets during the period of Late Heavy Bombardment (LHB). This idea was conformed to suggestions that the basin-forming projectile population is now extinct (Head et al., 2010; Werner, 2008). However, most recent estimates of one of the derivations of the Nice model suggest that only a minor portion of the projectiles is cometary (Bottke et al., 2012), which is in agreement with the estimates by Ivanov et al. (2002) of <15% of comet impacts for terrestrial planets. The latter point of view has been discussed in a set of papers (Hartmann, 1977; Hartmann et al., 2002; Ivanov, 2001, 2004; Ivanov et al., 2001, 2002) previously. The relative role of various projectile populations is still under investigation.

Despite the general success of the migration/Nice model (now often considered the mainstream), many detailed problems require further investigation. Because these problems are vital for estimates of the 'pre-LHB' bombardment rate, we briefly list selected recent papers. More detailed information on the progressing work on these models is summarized by Michel and Morbidelli (2013). In the basic gas drag/gravity migration model ('Nice I' according to Levison et al., 2011), the migration and the gravitational interaction action of planetesimals are explained by close encounters with giant planets as it was proposed by Fernandez and Ip (1984). One can assume that a few of these planetesimals could reach the inner solar system and raised the pre-LHB bombardment rate in the region of the proto-MAB and terrestrial planets (like modern ecliptic comets, e.g., Levison and Duncan, 1997). Internal problems of the model motivate the expansion of the model components resulting in conceptually new possibilities. Levison et al. (2011) proposed the 'Nice II' model with many innovations. Now, the planet migration is assumed to operate 'remotely' by the gravitational interaction with the outer planetesimal disk as a whole without the necessity of physically close encounters. In this case, the 'comet-like delivery chain' of planetesimals into the inner solar system (as in Levison and Duncan, 1997) may be much smaller than in the 'Nice I' model. Models of cometary flux through the MAB during LHB constrain the 'cometary' fraction but assume the 'death' of comets near to the Sun (Broz et al., 2013). The comprehensive discussion of the early bombardment flux is published by Fassett and Minton (2013).

The other issue of the planet migration models is the 'initial' state of planetary configuration at the time of vanishing of the gas envelope. In the 'Nice II' model, the configuration is assumed from the results of Morbidelli et al. (2007). Other models allow much stronger migration of Jupiter, approaching as close to the Sun as ~1.5 AU during the first 100 000 years (the 'Grand Tack' scenario by Walsh et al., 2011, 2012). The mechanism results in the appreciate mixing of 'dry' and 'wet' material in the inner solar system and is able to facilitate the proper Mars mass in the modeling (see the review by Brasser, 2013).

Another issue discussed is the fine details of the 'Nice event' (the resonance of Jupiter and Saturn). The model mechanism of the LHB is the resonance sweeping through the pre-LHB asteroid belt, which results in the redirection of small bodies toward planet-crossing orbits (Minton and Malhotra, 2009, 2010). The original estimates of the LHB duration and intensity have been found adequate for ~0.5 My characteristic time of the Nice event. However, the Nice model's typical characteristic time is ~5 My (Morbidelli et al., 2010). The reconciliation is proposed in the more complex Nice event scenario, which demands a third giant planet (Uranus or Neptune) to make a close encounter with Jupiter and Saturn, accelerating the Nice event ('jumping-Jupiter' model; Fassett and Minton, 2013; Morbidelli et al., 2010). Additional (now missing) planets are often discussed in the planet formation models to tune the outcome to the real solar system (Brasser et al., 2013; Nesvorný and Morbidelli, 2012).

For satellites of giant planets, comets seem to be most probably the main bombardment population (Zahnle et al., 1998, 2003). However, estimation of absolute impact fluxes for the outer solar system stays an extremely difficult problem at present levels of knowledge, since interplanetary source populations are poorly known. Cratering on the icy satellites of the outer planets is beyond the focus of this chapter. In

general, numerical models of the solar system evolution (n-body simulations) strongly depend on the initial conditions and thus cannot be considered as model-independent evidence.

The authors point out that with the conceptual Nice model, impact flux estimates of absolute ages for the outer solar system could be in the future calibrated using the onset and duration of the LHB period to postulate a solar system-wide event. However, the planet-crossing bodies that might hit or have hit Jupiter and Saturn satellites are still unexplored so that statistics are very poor regarding the size–frequency distribution (SFD) at small sizes. Dynamical methods are compromised because there may be various poorly studied resonant effects throwing objects onto Planetary crossing orbits (PCOs) (as with near-Earth objects (NEOs) from MAB resonances), and the parent populations such as KBOs, Centaurs, and Trojan swarms are not fully characterized. Surfaces such as on Europa, Titan, and parts of Enceladus, with notably small crater densities, are believed to have geologically young absolute ages, indicating ongoing heating and geologic activity, but we believe the error bars on ages must be much larger than the factors 2–4 that have been quoted for Mars. A more complete discussion of the impact cratering and cratering rate on outer planet's satellites is presented, for example, by Kirchoff and Schenk (2010), Schenk et al. (2004), and Zahnle et al. (1998, 2003). A discussion of the bombardment history in the context of the LHB concept is published by Charnoz et al. (2009) and Barr and Canup (2010).

10.10.3 Impact Craters: Morphology

Impact craters are circular depressions with elevated rims (Figure 1) created by the collision of celestial bodies. The impact crater formation occurs as a chain of subsequent processes: it starts with a high-velocity impact and ends with ejected material deposition and crater cavity adjustment to the new equilibrium of fragmented rocks in the planetary gravity field. First observations of impact craters have been done with telescopic observations of the Earth's Moon starting with Galileo Galilei. Impact craters were eventually recognized on the Earth's surface, particularly in the 1950s and 1960s (Dence, 1965; French and Short, 1968). With the beginning of space flights, the scientific interest in craters has grown enormously. Both imaging of extraterrestrial planetary bodies and studying of terrestrial impact craters advanced our understanding of the cratering process. In terms of understanding times and rates of geologic processes on planets, we are fortunate that nature *stamps* sharp, symmetrical circular structures on planetary surfaces. If the crater formation rate is understood, we can estimate ages from crater frequencies and interpret the nature and rates of geologic activity, such as erosion and deposition processes, from morphological degradation and losses of craters.

The close-up imaging of the Moon has revealed the fundamental difference of the impact crater population from other (e.g., volcanic) crater structures, namely, the wide spectrum of crater sizes, continuing from microns to thousands of kilometers.

The size–morphology progression ranges from small, simple bowl-shaped craters, to craters with central peaks and terraces at the walls (an example is shown in Figure 1), up to basins with inner rings and, on some bodies, exterior scarps that encircle the impact site. A collection of papers in the book *Impact Cratering: Processes and Products* (Osinski and Pierazzo, 2013) gives an extended introduction to the impact cratering topic.

The morphological diversity of craters seems not a direct result of the crater excavation process, but develops only as the explosion cavity evolves just after most of the material has been ejected from the initial cavity. The initial product of crater excavation is considered to be a circular, bowl-shaped *transient* cavity with a depth/diameter ratio between 1/4 and 1/3. This transient crater then undergoes different degrees of modification as a result of gravitational instability and collapse of the crater walls and the rebound of the crater floor. The final crater morphology depends on the gravitational acceleration, density, and condition of the surface material of the target planet. This affords both the opportunity of using impact crater morphology to learn about conditions in the target and the challenge of understanding how different conditions affect the course of crater collapse (Melosh and Ivanov, 1999).

Dence (1965) first classified terrestrial impact craters as either simple or complex in structure. This classification,

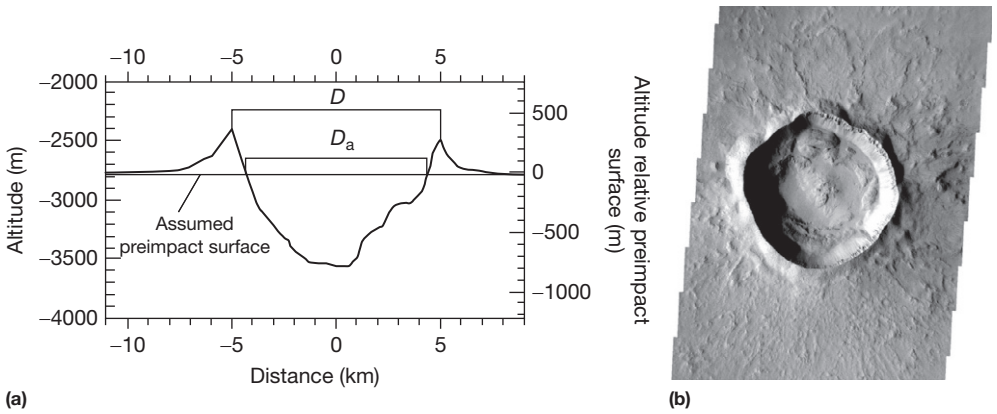


Figure 1 The profile of a typical fresh impact crater (exemplified with the 10 km diameter Martian crater Zunil, profiled by MOLA track, a). This north-to-south profile does not reveal the small central mound visible in high-resolution image (b, THEMIS image V09818024). The figure introduces the often used rim crater diameter, D, and the apparent crater diameter, D_a, measured at the preimpact surface level.

based on the study of craters exposed by erosion, also seems to apply to the morphology of fresh craters revealed by images of the surfaces of other planets and satellites. Data sets on terrestrial and planetary craters are highly complementary: Due to erosion, geophysical investigations, and direct drilling, the subsurface structure of many terrestrial craters can be explored in detail, while the surface morphology is often displayed in exquisite detail by images of fresh craters on the surfaces of other planets or satellites (for a brief review, see Wright et al., 2013).

Simple craters are circular, bowl-shaped depressions with raised rims and approximately parabolic interior profiles. The rim-to-floor depths of a large sample of simple craters on the Moon are roughly 1/5 of the rim-to-rim diameter (Pike, 1977a). Simple craters have a true floor, covered with a breccia lens up to a half of the *true crater* depth. Drilling results and seismic investigations prove that this lens is created by fragmented material avalanched from the transient crater walls down into the crater, forming the *apparent* crater floor (Grieve et al., 1977; Shoemaker, 1963). The volume of the breccia lens is roughly one-half of the volume of the crater itself (Grieve and Garvin, 1984; Settle, 1980).

Complex craters possess a more complicated structure than simple craters (for a recent review, see Kenkmann et al., 2013). On the Earth, they exhibit central structural uplifts, rim synclines, and outer concentric zones of mainly normal faulting. Images from spacecrafts show extraterrestrial craters with single or multiple central peaks, flat inner floors, and terraced rims (**Figure 1**). The depths of complex craters increase with increasing diameter but much more slowly than the depths of simple craters (Pike, 1977a,b).

The transition between simple and complex craters occurs over a relatively narrow diameter range on all solar system bodies and seems to scale as the inverse power of the surface gravity (e.g., Hartmann, 1972). Thus, the simple–complex transition (Pike, 1980) is relatively well determined on the Moon at about 15 km in diameter. On Mercury and Mars, it is about 11 and 7 km in diameter, correspondingly, and 2–4 km in diameter on Earth (depending on whether the crater forms in sedimentary or crystalline rocks). On icy satellites, the transition depends on the low-temperature ice strength and the presence or absence of a sub-ice liquid mantle (Moore et al., 2001; Schenk, 2002; Schenk and Ridolfi, 2002; Schenk et al., 2004). Hartmann and Wood (1971) described a broadening of central peaks into ring of peaks, for which they coined the term *peak ring*. Peak rings possibly mark a transition to fully configured multiring basins.

Impact basins are the largest impact structures. On some bodies, they exhibit many rings, being named as multiring basins. The classic basin example is Orientale on the Moon's western limb (eastern limb in terms of sky coordinates used by early telescopic observers). First recognized on rectified lunar photographs in 1962 (Hartmann and Kuiper, 1962), multiring basins appear to form as a tectonic response of the lithosphere to the crater created by the impact (McKinnon and Melosh, 1980; Melosh and McKinnon, 1978). A review of morphology and occurrence of basins is published by Spudis (1993). The data from recent space planetary missions resulted in the enlarged volume of data about basins and their deep structure on the Moon (Crosby and McKenzie, 2005; Fassett et al., 2012a; Pieters et al., 1997; Potts and von Frese, 2003a,b; Sugano and Heki, 2004; Wieczorek and Phillips, 1999) and Mars (Hiesinger and Head, 2002; Neumann et al., 2004; Potts et al., 2004; Zuber et al., 2000). New data about Mercury have also been published (Andre and Watters, 2006; Andre et al., 2005; Baker and Head, 2013; Fassett et al., 2012b).

10.10.4 Basics of Impact Cratering Processes

Craters form in various shapes and under different conditions on each of the solid planetary bodies in the solar system. If the cratering process is well enough understood, craters make up a unique sensing system for understanding the evolution of the solar system. With a good set of crater scaling laws, relating crater and projectile SFDs, the role of the underlying geology can be derived. The key parameters are not only the projectile size and velocity but also material properties. Theoretical solutions make use of the analogy between nuclear explosions and impact crater formation in their initial phases by a point source approximation (review by Holsapple, 1993). Thus, it allows scaling the cratering process by power-law forms and the impact energy relates through a coupling parameter to the *transient* crater cavity size. The later modification stages are crater size-dependent and need to be considered separately to define the final crater size. Instructive introductory reviews were recently published by Melosh (2013) and Collins et al. (2013).

10.10.4.1 Principal Stages of Impact Crater Formation

The association of high-pressure phases of quartz, coesite and stishovite, is a reliable mineralogical criterion for an impact crater origin. It also suggests that the impact process as a whole can be described in analogy to an explosion, in which the initial kinetic energy of the projectile does work on the target to create a visible depression (the crater), to shock-metamorphose the target material, and to heat both projectile and target. The sequence of events in a high-velocity impact may be presented as a set of several sequential stages. Each of these stages is dominated by a specific set of principal physical and mechanical processes. It is convention to distinguish three main stages of an impact event (Gault et al., 1968; Melosh, 1989). These stages are (1) contact and shock compression, (2) transient crater cavity growth by crater material ejection, and, finally, (3) transient crater cavity modification (slumping or collapse). These stages are not strictly separated, but are used for convenience in the analysis of the impact process because they highlight the dominant mechanisms.

During the first stage of *contact and shock compression*, the energy released by the deceleration of the projectile results in the formation and propagation of shock waves away from the point of impact. Peak shock pressure is of the order of the stagnation pressure ρU^2, where ρ is the smaller of either the projectile or target density and U is the impactor velocity. At velocities above a few km s^{-1}, the pressure greatly exceeds the strengths of rocks (a few GPa at high pressure) and the process can be treated hydrodynamically. Shock waves propagate out of the point of impact, decay, transform to stress waves and, at larger distances, to the elastic (or seismic) waves.

After release from high shock pressure, accelerated target material moves out of the impact point, and *excavation and opening the transient crater cavity of a growing impact crater* follows. Studies of the crustal thickness beneath large lunar basins support the hypothesis of *proportional scaling* for the dimensions of the transient crater cavity up to craters at least 500 km in diameter (Wieczorek and Phillips, 1999). In many numerical model simulations, the floor of the *transient* crater achieves its maximum depth and begins to rise into a central peak while the diameter is still increasing (Melosh, 1989). The extent of the subsequent modification of the transient crater cavity depends on the size of the crater. The hiatus between the point in time when the maximum crater depth is reached and the point in time when the radial excavation flow has ceased is more pronounced the larger the crater and the more important the effect of gravity. In small craters, for example, at laboratory scale at which most scaling laws were derived, the definition of the transient crater is trivial as it is the final crater dimension.

Details of the *transient crater cavity modification* by wall slumping for simple craters are presented by Grieve and Garvin (1984). A full-scale review of the mechanics of the transient crater cavity collapse in the case of complex crater formation is given by Melosh and Ivanov (1999). The general overview of complex crater formation presented by Melosh (1989) is still the best introduction. Some particular issues are updated in more recent papers about numerical modeling of impact craters (Collins et al., 2002, 2008; Ivanov, 2005; Morgan et al., 2000; O'Keefe and Ahrens, 1993, 1999; O'Keefe et al., 2001; Senft and Stewart, 2008).

10.10.4.2 Impact Crater Scaling Laws

Scaling laws relevant to impact cratering describe a set of rules, both experimentally derived and theoretically assumed, and formulated to establish relationships between properties and parameters of an impactor (projectile), the planetary surface (target), and the resulting impact crater under conditions of different planetary bodies, such as gravity and specific material properties. Originally, scaling laws were formulated to make a theoretical interpolation of laboratory experimental data, mostly for metal targets (Kinslow, 1970), and to extrapolate the results of impact experiments at velocities achieved at laboratory-scale to a natural impact velocity range. Later, the focus changed to rock and soil targets when planetary scientists tried to estimate the size of projectiles that created craters on planetary bodies. These efforts on rock and soil targets and explosion cratering studies were collected by Roddy et al. (1977). While for metal targets, the shear strength of metals dominantly limits the size of the growing transient cavity, for natural planetary targets, where crater diameter ranges from less than millimeters to more than 1000 km, one observes a transition from strength-dominated formation of small craters to gravity-dominated formation of large craters. *Large* and *small* crater diameters depend on the intrinsic target strength, which adjusts in a complicated way during the cratering process. This adjustment is influenced by the brittle damage of the intact rocks, additional fragmentation of real fractured rocks, mechanical dispersion of shock energy in sands and regoliths, temporary decrease of friction, and thermal softening, when the rock strength vanishes near to the melting point. For this reason, it is convenient to refer to *effective* rock strength, usually implemented as one parameter in the available theoretical models. However, this parameter, in reality, is highly variable in time and space during the cratering process.

Several approaches to scale cratering effects for a variety of impact velocities, target strength parameters, and an impact angle exist. They are used to scale crater size observed on different planets to find the crater-forming projectile size, which is then used to compare the similarities and differences of cratering rates. The reliability of this comparison defines the accuracy of interplanetary crater-retention age recalculations. Scaling relations seek to find analytic descriptions, which relate input parameters at ranges that can cover several orders of magnitudes and the dimensions of the final parameter without the simulation of the entire physical process. Early theoretical studies on self-similar solutions of the second kind (Zeldovich, 1956) have resulted in a concept, later named 'the late-stage equivalence' (Dienes and Walsh, 1970): The impact cratering process at distances much larger than the projectile size is similar for all impacts in the same target, if the value of the so-called coupling parameter (some combination of the mass and velocity of the projectile) is constant. There exist several approaches of scaling impact crater formation.

10.10.4.2.1 The concept of coupling

The central point of the coupling-parameter concept is that the cratering efficiency of one impact will be the same for all projectiles with the same coupling-parameter value. Because the details of the energy release at the source point do not affect the later process significantly, Dienes and Walsh (1970) had proposed a coupling parameter to scale the shock pressure decay under the point of impact in the form of a characteristic length scale, L, for a given impactor and, thus, a measure for the penetration depth:

$$L = D_p(\delta/\rho)^{\nu}(U/c_b)^{\mu} \qquad [1]$$

where D_p is the projectile diameter, δ/ρ is the projectile-to-target density ratio, U is the impact velocity, and c_b is the bulk velocity of the target material, defined as $c_b = (K/\rho)^{0.5}$, where K is the bulk compression modulus. ν and μ are dimensionless exponents, which are derived from experiments. For the density-ratio exponent, the value of $\nu = 1/3$ gives a good approximation for available data. The velocity exponent μ is bounded by theoretical extreme values for the velocity exponent related either to a maximum when scaled by the energy ($\mu = 2/3$) or to a minimum when scaled by the momentum ($\mu = 1/3$) of the impactor. For the same projectile density, a simple *energy scaling* demands that the value $D_p U^{2/3} = \text{const.}$, while experiments and numerical modeling reveal an *equivalence of mechanical action* (e.g., the same crater dimensions) for projectiles with $D_p U^{\mu} = \text{const.}$ only for velocity exponents $\mu < 2/3$, for example, $\mu \approx 0.55$–0.58 for competent nonporous targets and $\mu \sim 0.45$ for porous targets like dry sand. The fundamental reason for the inequality $\mu < 2/3$ is that at higher impact velocity, a small amount of material near the impact point is overheated and ejected, and that decreases the amount of energy available for the crater formation at later stages. The lower theoretical limit resulting from *momentum scaling* demands that $\mu > 1/3$ (Holsapple, 1993; Holsapple and

Schmidt, 1982). The transition between strength- and gravity-dominated regimes depending on projectile size is not yet considered. The transition is gradual and becomes important in the late stage of the transient crater cavity growth.

The scaling of transient crater cavity depth, z, below the impact point with the length scale L allows us to present the shock pressure, P, decay in the form

$$P/\rho c_b^2 = f_p(z/L) \quad [2]$$

where f_p is the universal function for a wide range of impact conditions. Experimental data for metal targets are given by Dienes and Walsh (1970). Progress in numerical modeling techniques promoted the investigation of shock pressure decay in *geomaterials* such as rocks. Almost all rock components typical for planetary crusts pass through a solid–solid phase transition under compression (quartz–coesite–stishovite transition is a typical example) and make the scaling of impact cratering in rock more complicated than for metal targets. However, available numerically modeled data show reasonably good convergence to a general shock pressure decay function, as given in dimensionless form of eqn [1]. For target material with a simple equation of state, Ivanov (1981) showed that the shock pressure decay may be expressed as

$$P/\rho c_b^2 = (3 \pm 1)(z/L)^{-2.5} \quad [3]$$

with the best choice of $\mu = 0.58$ recommended by Dienes and Walsh (1970). Pressure decay with distance from the impact point in geomaterials may happen on a shorter length scale than, for example, in metals, because the solid–solid phase transition and initial porosity require additional energy, which dissipates by additional residual heating of the shocked material.

A short note should be added about the widely cited concept of a so-called isobaric core. Numerical modeling by Pierazzo and Melosh (1999) revealed that the geometry of shock pressure isobars may be represented by a truncated sphere below the point of impact. In the limits of a certain isobar near the projectile contact point, the maximum pressure is almost constant and close to the initial contact pressure. This zone is cited as an 'isobaric core.' The reader should be aware that inside the 'isobaric core,' the shock pressure is about the same, indeed, but the contours of equal shock pressures do not coincide with the surface of the shock wave – the same shock pressure inside the 'isobaric core' is reached at various instants of time close to the centerline and far from it. The effect is clearly shown by Melosh (2013) and Figure 3.1 therein. The instant shock front surface is much closer to a hemisphere with the center in the contact point than a truncated spherical surface.

Holsapple and Schmidt (1980), Holsapple and Schmidt (1982), Holsapple and Schmidt (1987), and Schmidt and Housen (1987) had found a set of parameters to scale cratering dimensions, not explicitly considering target material compressibility yet. The coupling (parameter) of impact energy or momentum into the target depends on the specific target (rock) properties (e.g., solid–solid phase transitions, porosity, and effective strength). Thus, for example, in numerical models, it stays challenging to incorporate the momentum or energy transfer correctly, producing the appropriate shock-wave propagation. The shape of the shock front depends on the sets of coupling parameters, but it has been proven to be useful in many applications of *high-velocity* impacts when the impact velocity is larger than the sound velocity of the target and the projectile is not infinitely strong. For *low-velocity* impacts, which are subsonic relative to the target material (sound velocity), and/or for deeply penetrated projectile, the coupling-parameter function may be different, particularly if it depends on impact velocity not as an exponential function. For example, low-velocity impacts in sand are well described with a coupling-parameter proportional to a linear combination of U and U^2 terms (Mizutani et al., 1983). In the succeeding text, we concentrate on high-velocity impacts.

10.10.4.2.2 Strength and gravity scaling

For fixed values of the coupling parameter, the kinetic energy of the excavation flow accounts for either mostly plastic work against the strength of target material (*strength regime*) or mostly potential energy of the growing cavity in the gravity field (*gravity regime*). In the former case, the energy consumed during the transient cavity growth to a characteristic size a (the linear scale may be described by crater depth, radius, or the cube root of the volume, $V^{1/3}$) is proportional to Ya^3 (the multiplication of a transient crater volume, proportional to a^3 by the effective strength Y), while in the latter case, the consumed energy is proportional to $\rho g a^4$ (the multiplication of the mass extruded from the transient cavity, proportional to ρa^3, by the gravity acceleration, g, and the effective height of the mass uplift, proportional to a). Craters formed in the gravity regime need increasingly more energy to make a larger crater. These large craters are relatively small with respect to their projectile and shock melt-zone size. This implies that the largest impact basins on terrestrial planets may have transient cavities only a few times larger than the projectile size and the exact proportions depend on the assumed impact velocity. In the intermediate-sized crater range, both gravity and strength energy consumption mechanisms operate simultaneously.

There are many ways to express the transition from strength-to-gravity regimes. The basic experimentally derived scaling laws (Schmidt and Housen, 1987) use the π-theorem (Buckingham, 1914). For the case of crater scaling, the π-scaling is described in detail by Melosh (1989) (for a review, see Holsapple, 1993):

$$\pi_D = K\pi_2^{-\alpha} \quad [4]$$

where π_D is the scaled transient crater diameter D_a,

$$\pi_D = D_a(\rho/m)^{1/3} \quad [5]$$

The gravity-scaled projectile size (for a sphere of a diameter D_p equivalent mass) in the gravity-dominated regime is given by

$$\pi_2 = (1.61 g D_p)/U^2 \quad [6]$$

and in the strength-dominated regime, it is

$$\pi_2 = Y[\rho U^2] \quad [7]$$

where Y is an effective strength of the target material, keeping in mind that this property is highly variable during the cratering process in space and time. The use of a single value can only refer to the averaged value.

Here, D_a is the transient apparent crater diameter, ρ is the target density, m is the projectile mass, U is the impact velocity, g is the gravitational acceleration, and D_p is the projectile diameter. Schmidt and Housen (1987) determined coefficients in eqn [4] for various materials from experimental data, with coefficients for some typical rock types such as for porous (e.g., dry sand) targets $K=1.68$ and $\alpha=0.17$ and for nonporous rocks $K=1.6$ and $\alpha=0.22$. The parameters for nonporous rocks can be experimentally determined only for the strength regime, and the values given here are not necessarily applicable for the gravity regime scaling.

The elaboration of the hydrocode-friendly model of porosity (Wünnemann et al., 2006) and the extensive modeling of vertical and oblique impacts (see succeeding text) resulted in the appropriate change of terminology. What we previously named *nonporous* scaling law is more appropriate to name "*nonporous, zero friction* scaling law. Note that Schmidt and Housen (1987) calibrated this law by experiments in water-saturated sand (the filling of pores with water sufficiently decreases friction). The formerly *porous* scaling law calibrated by dry sand experiments appears to have its different exponent in eqn [4] mostly due to dry friction. It is more appropriate to name it *dry friction* scaling law (porosity decreases the coupling efficiency but keeps the α value close to 0.17 for a wide range of porosities – at least in the available numerical models; Nowka et al., 2010; Wünnemann et al., 2011, 2012).

Figure 2(a) and **2(b)** illustrates the scaling for dry and wet sand (*dry friction* and *nonporous, zero friction*) rocks for comparison and the experimental data used to define scaling constants. Porous rocks, dry sand, and lunar regolith experimental data are shown in comparison. The scaling law represents the data well (**Figure 2(a)**). Nonporous rock scaling laws are compared to wet sand data. Schmidt and Housen (1987) observed a mismatch between their experimental crater size values for wet sand targets and an exponential extrapolation for the pure gravity scaling, because the experiments covered the strength–gravity transition region. The experimental laboratory data can be expanded by numerical modeling results, in which both the *transient* crater cavity diameter (all scaling laws refer to the transient crater cavity only) and the *final* rim diameter have been recorded. In **Figure 2(b)**, numerical results for terrestrial craters with rim diameters from ~5 to ~200 km in layered targets (sediments over granitic crust over dunitic mantle) form a natural continuation of the high-G experimental data for wet sand, but with a shallower slope (closer to a slope for porous targets) in π_D–π_2 coordinates. (In centrifuge experiments, G is the artificial gravity value normalized as the ratio of the centripetal acceleration to the normal terrestrial free fall gravity acceleration $g=9.81\ \mathrm{ms}^{-2}$. These experiments make use of the proportionality between gravitational acceleration and projectile energy, so that impact experiments in a centrifuge can extend the transient crater cavity range to larger diameters when normal terrestrial free fall gravity acceleration is considered.) The reason for this behavior may be the presence of dry friction in numerical models, not well discussed in early nonporous scaling law derivation (Schmidt and Housen, 1987). Both experimental (Schmidt, 1980) and numerical modeling (Wünnemann et al., 2006) data show that porosity and dry friction in targets should be decoupled in an updated version of scaling technique now under construction.

The first steps toward updated scaling laws have been done recently (Elbeshausen et al., 2009; Nowka et al., 2010; Wünnemann et al., 2011, 2012). These authors published a series of numerical models where the transient crater dimensions were systematically calculated for various values of the dry friction coefficient and porosity of the target material (**Figure 2(c)** and **2(d)**). Most recent experimental work on porous, wet, and dry sandstone suggests a complex relation of porosity and pore space saturation to influence the cratering process, thus a challenge to estimate the relation between the transient crater cavity and final crater diameter. In the strength regime, a coincidence of these two measures is assumed (Poelchau et al., 2013).

The π-theorem relations may be rewritten in the following familiar form, for nonporous, zero friction rocks:

$$D_{tc} = 1.16\,(\delta/\rho)^{1/3} D_p^{0.78} (U)^{0.44} g^{-0.22} \quad [8]$$

and for dry friction (dry sand) material:

$$D_{tc} = 1.25\,(\delta/\rho)^{1/3} D_p^{0.83} (U)^{0.34} g^{-0.17} \quad [9]$$

Relations presented earlier in the text are valid for the gravity regime of cratering. The procedure of how to approximately include the strength–gravity transition, inspired by Holsapple and Schmidt (1979), is described in Neukum and Ivanov (1994) and Ivanov et al. (2001). Arithmetically, the engineering equation for nonporous, zero friction rocks is transformed to

$$\left(D_{tc}/D_p\right)^{0.78} = 1.16(\delta/\rho)^{1/3} (U)^{0.44} (gD_{tc})^{-0.22} \quad [10]$$

and consequently to

$$D_{tc}/D_p = \left[1.16(\delta/\rho)^{1/3} (U)^{0.44} (gD^{tc})^{-0.22}\right]^{1/0.78} \quad [11]$$

Finally, all numerical values for δ, ρ, U, are combined into one coefficient C_D,

$$D_{tc}/D_p = C_D/(gD_{tc})^{0.282} \quad [12a]$$

and, correspondingly, for 'dry sand' (dry friction) scaling (eqn [9])

$$D_{tc}/D_p = C_D/(gD_{tc})^{0.205} \quad [12b]$$

In the strength regime, the ratio of D_{tc}/D_p is a constant. To make a smooth transition from the strength-to-gravity regime, we assume that

$$D_{tc}/D_p = C_D/\left[g\left(D_{sg} + D_{tc}\right)\right]^{0.282}$$
for nonporous, zero friction rocks, and [13a]

$$D_{tc}/D_p = C_D/\left[g\left(D_{sg} + D_{tc}\right)\right]^{0.205}$$
for dry friction (dry sand) material [13b]

where D_{sg} is the effective boundary between gravity- and strength-dominated regimes of crater formation. For the discussion on the actual D_{sg} value for terrestrial rocks, see Holsapple and Schmidt (1979), Croft (1981), and Neukum and Ivanov (1994).

Here, we have a smooth approximation of the transition to strength craters for small craters $D_{sg} \ll D_{tc}$, and the D_{tc}/D_p ratio is constant for a given C_D value:

$$D_{tc}/D_p = C_D/\left[gD_{sg}\right]^{0.282}$$
for nonporous, zero friction rocks, and [14a]

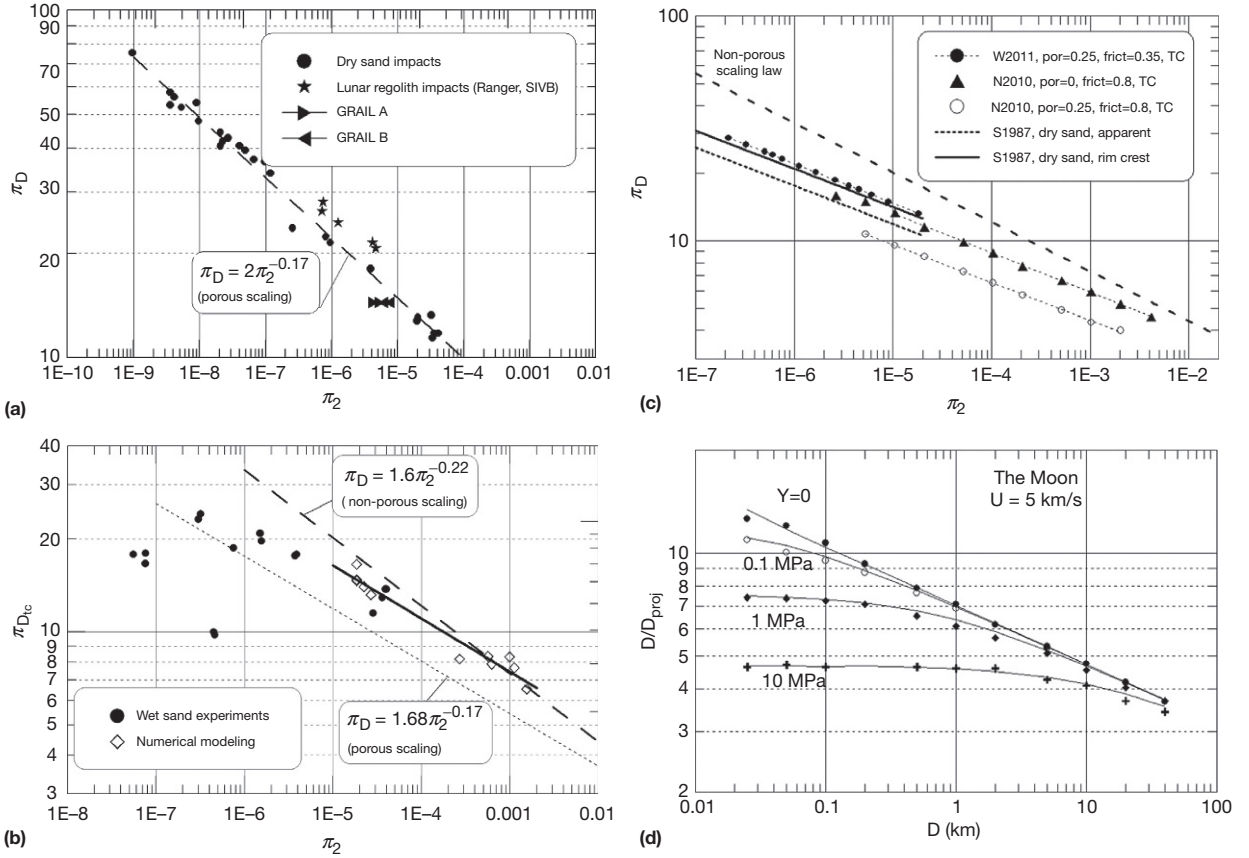


Figure 2 (a) π-Scaling for dry sand ('porous dry friction rock') for the rim crater diameter. Laboratory experimental points (Housen and Holsapple, 2003; Schmidt, 1987) are compared with artificial lunar craters created by Ranger SIVB, and GRAIL spacecraft impacts (Moore, 1972; Witaker, 1972).(Lunar crater diameters are measured on recent LRO image data (http://lroc.sese.asu.edu). GRAIL craters are formed at a steep slope of ~25° at small velocities of 1.6 km s^{-1}. The crater and topographic data are provided at http://lroc.sese.asu.edu/news/index.php?/archives/719-Impact!.html. For the assumed crater diameter of 6 m, the uncertainty in the average projectile density in the range from 130 to 260 kg m^{-3} is shown. Being smallest of all artificial impact craters, GRAIL craters may manifest the influence of the regolith cohesion, resulting in downshifted position of π_{Dtc} values.) (b) Scaled transient crater radius π_{Dtc} versus scaled projectile size, π_2, from experimental Boeing Lab data (Schmidt and Housen, 1987) for wet sand (black circles), and from numerical modeling (Ivanov, 2005; Ivanov and Stoffler, 2005; Joeleht et al., 2005; Stöffler et al., 2004) as open diamonds, transient cavity is defined as in (Wünnemann et al., 2006). Numerical modeling data appear as a natural continuation of experimental data, following however an exponent of −0.17, similar to experimental exponent found for dry sand, but not as steep as the assumed 'nonporous' value of −0.22 (Schmidt and Housen, 1987). It is unclear whether finite friction in the numerical model is the cause for this observation. (c) Numerical modeling results for porous targets with dry friction in comparison with experimental data for dry sand (Schmidt, 1987; solid black line is for rim crest diameter, dashed line is for the apparent diameter; S1987 in the legend). Black circles are model data for the transient cavity from Wünnemann et al., 2011, for 25% porosity and friction coefficient of 0.35. Triangles (zero porosity) and open circles (25% porosity) are for scaled transient cavity diameters in a material with friction coefficient of 0.8 (Nowka et al., 2010). Upper dashed line is for the 'nonporous, zero friction' scaling law (scaled eqn [4] and dimensional eqn [10]). (d) Numerical models for transient cavity dimensions in a nonporous material with various strengths (cohesion) published by Nowka et al. (2010) in dimensional coordinates (lunar gravity, impact velocity of 5 km s^{-1}).

$$D_{tc}/D_p = C_D/[gD_{sg}]^{0.205}$$

for dry friction (dry sand) material [14b]

The values of $1/[gD_{sg}]^{0.282}$ and $1/[gD_{sg}]^{0.205}$ represent the effective target strength parameter, so that the smaller the strength, the larger the crater and the smaller the transitional diameter between craters formed in strength and gravity regimes.

Preliminary (transient cavity only) modeling by Nowka et al. (2010) and Wünnemann et al. (2011) illustrates how target porosity decreases the crater efficiency of an impact (**Figure 2(c)**). For example, porosity of 25 vol% decreases the π_D value by a factor of 1.6 (with a proper account for smaller density of the porous material). One more effect to study in the near future is the thermal softening of geomaterials with the dry friction. Heating of materials due to the shock compression/decompression heating and the plastic work may decrease the friction coefficient. It results in the gradual change from the 'dry friction' scaling law (eqn [14b]) to the 'no friction–gravity-dominated' scaling law (eqn [14a]). Preliminary modeling (Ivanov and Kamyshenkov, 2012) illustrates the transition. Previously, the idea of the thermal softening was proposed for the yield strength (cohesion) decrease only (O'Keefe and Ahrens, 1993, 1999).

Figure 2(d) illustrates the strength–gravity transition in the model target material with zero porosity, constant friction coefficient, and variable strength (cohesion) computed by Nowka et al. (2010). Theoretical curves are fitted to individual model data with eqn [13b]. The comparison of the cohesion value, Y, in the iSALE code and the characteristic strength/gravity transition diameter D_{sg} gives an approximate rule for the lunar gravity: $D_{sg}(\text{m}) \cong 5\,Y(\text{kPa})$. Note that 1 bar cohesion is equivalent to 100 kPa.

More studies of nonuniform targets (e.g., where porosity decreases with depth) are needed to model impact craters on asteroids (8–10% of a bulk porosity of Vesta; Toplis et al., 2013) and the Moon (Wieczorek et al., 2013), where porosity is proven to exist in the whole crust and upper mantle. Some preliminary models of a nonuniform Moon are published by Wünnemann et al. (2012).

The transient crater cavity diameter is proportional to the rim crater diameter as $D \approx (1.25 \pm 0.05) D_{tc}$ for simple craters, and, according to Croft's model (Croft, 1985),

$$D_{tc} = D_{sc}^{0.15} D^{0.85} \qquad [15a]$$

or, vice versa,

$$D = D_{sc}^{-0.18} D_{tc}^{1.18} \qquad [15b]$$

for $D > D_{sc}$, where D_{sc} is the simple-to-complex crater transition diameter. One should note that the simple-to-complex transition is poorly investigated and the recipe mentioned earlier in the text is the simplest but an internally consistent one. Equations [15a] and [15b] were obtained from the geologic reconstruction of partially eroded terrestrial craters. Numerical modeling gives an opportunity to get similar relations directly by the comparison of transient cavity diameter and the final rim diameter of a collapsed crater (Figure 3). The exponent in eqn [15b] is close to model values (Figure 3) and proves the general validity of Croft's model. However, a final conclusion is only reached with further natural and model data reconciliation.

For other planetary bodies, the D_{sc} value may be assumed from crater morphology transition (Pike, 1980). Numerical modeling of the terrestrial impact crater formation extended to the end of target material motion (see references in the caption for Figure 2) allows a final approximation of the relationship between projectile diameters, D_p (in km); impact velocities, U (in km s^{-1}); and the final rim crater diameter, D (in km), for a vertical impact as

$$D = 4\left(D_p U^{0.58}\right)^{0.91} \qquad [16]$$

valid for the diameter range shown in Figure 3. The general behavior of the D/D_p ratio for craters in lunar nonporous rocks has been presented in Neukum and Ivanov (1994) and Ivanov et al. (2001). The general D/D_p dependence on the final crater diameter for lunar craters is shown in Figure 4, constructed with the impact velocity of 18 km s^{-1}, lunar regolith density of 1900 kg m^{-3}, and projectile and bedrock density of 2700 kg m^{-3}. Equation [16] and the scaling relation shown in Figure 4 will be used to compare SFD of projectiles and impact craters on the Earth and the Moon (see Section 10.10.6.2).

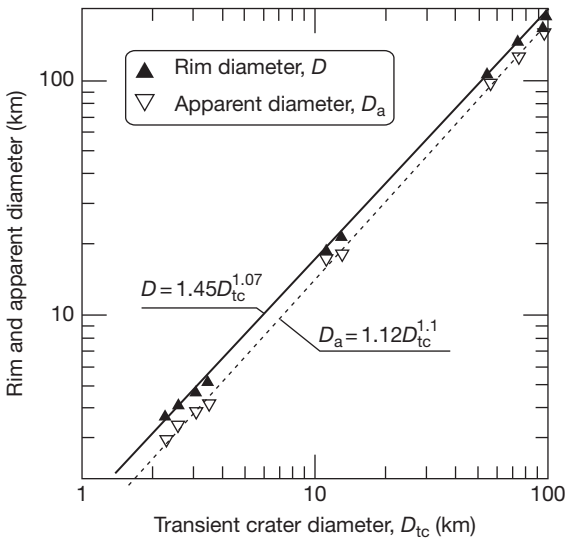


Figure 3 The relation between transient crater diameter and both rim crest and apparent final complex crater diameters from direct numerical modeling of terrestrial craters with diameters from ~4 to ~200 km (see references in Figure 2(b)).

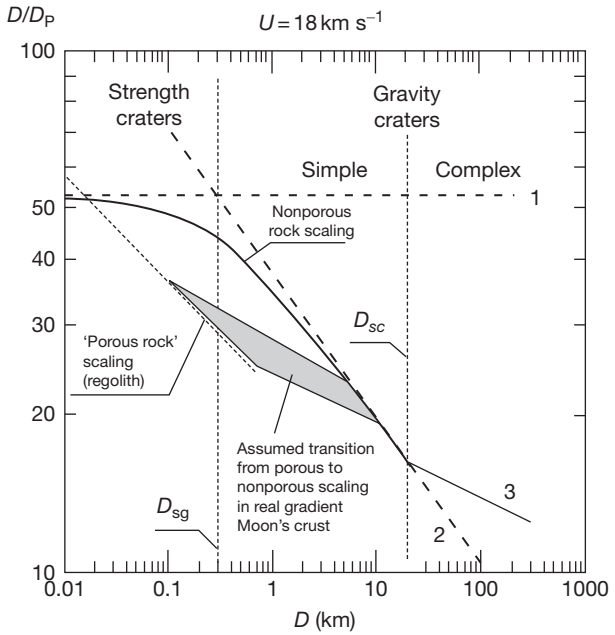


Figure 4 An example of the lunar crater scaling. The upper layer of the Moon is constructed of porous regolith, and for an assumed impact velocity of 18 km s^{-1}, the crater/projectile diameter ratio is of the same value as for the strength scaling for craters in nonporous bedrocks. For craters in nonporous bedrocks, the D/D_p ratio gently declines from strength scaling (D/D_p = const., dashed line 1) toward the gravity scaling (dashed line 2), but at $D \sim 15$ km, craters undergo widening due to the gravity collapse. For realistically layered lunar crust with a gradual transition for a relatively porous megaregolith to the solid rocks, one may assume a gradual transition from the porous to nonporous scaling for D between ~0.5 and ~10 km. Future studies of cratering in variably porous lunar crust should clarify the presented qualitative scheme.

10.10.4.2.3 Scaling of the impact angle

Vertical impacts of celestial bodies into planetary surfaces are suitable for computational and experimental studies. However, both vertical and grazing impacts have a very low probability to occur. It was shown for nongravitating spherical bodies (Gilbert, 1893) and for spherical planets with their natural gravitational attraction (Shoemaker, 1962) that the probability of impacts of a random flux of projectiles from infinity at various angles has a domical-shaped angle distribution with a maximum at 45°:

$$f_\phi = 2\sin\phi\cos\phi\,\mathrm{d}\phi \qquad [17]$$

where f_ϕ is the probability of impact in the interval of angles from ϕ to $\phi+\mathrm{d}\phi$. The convention is to measure the angle of impact from the horizon is $\phi=0°$ for grazing impacts and $\phi=90°$ for vertical impacts. Equation [17] is considered valid for planets and projectiles revolving around the Sun, where they form circular craters under most impact angles. Depending on the target properties and its *effective* strength, the crater becomes more elongated for impact angles below about 15° (Gault and Wedekind, 1978). This critical angle can vary depending on the crater efficiency (Collins et al., 2011; Elbeshausen et al., 2013). An oblique impact is less efficient than a vertical impact. Experimental modeling reveals, for example, that for the same projectile and the same impact velocity, the crater volume, V, decreases approximately as

$$V(\phi) = V(90°)\sin^\gamma\phi \qquad [18]$$

where $V(90°)$ is the crater volume for vertical impacts and the exponent γ is in the range from 1 to 2 depending on the target properties. Gault and Wedekind (1978) proposed γ between 1 and 2 for impacts in monolithic granite blocks and projectiles of aluminum, representing the strength regime, and $\gamma=1$ for impact in porous targets prepared of noncohesive dry sand and volcanic adhesive pumice, which is analog to the gravity regime. These results are widely cited despite the obvious difference in impact conditions: At a maximum impact velocity, v, of 7 km s^{-1}, impacts into sand and pumice are supersonic with v four to five times larger than the sonic speed in the target material, while for solid rocks, it occurs in subsonic to sonic regime, because the velocity v of 5–7 km s^{-1} is close to the longitudinal elastic wave velocity of solid rock. Pierazzo and Melosh (2000b) tested numerically the assumption that the experimental results of Gault and Wedekind (1978) can be approximated by assuming that the value of the velocity U in scaling laws for vertical impacts for an oblique impact may be replaced with $U\sin\phi$, representing the vertical component of a projectile velocity (Pierazzo and Melosh, 1999, 2000a,b). However, the real nature of oblique impact looks more complicated. The repetition of such experiments for monolithic rocks (Burchell and Whitehorn, 2003) confirms the validity of $\gamma=1$ similar to the porous target value obtained by Gault and Wedekind (1978). Preliminary last-stage numerical modeling (Artemieva and Ivanov, 2001; Ivanov and Artemieva, 2001, 2002) revealed that for a velocity of 20 km s^{-1}, the efficiency of oblique impacts decreases with $\sin\phi$. This is an observation that indicates a decrease less dramatically than suggested by the vertical velocity component paradigm. The issue deserves further investigation. As a tentative recipe, one can take into account that the 45° and steeper high-velocity impact (15 km s^{-1} and faster) may be as effective as the vertical impact. For low-velocity impacts (5–6 km s^{-1}), preliminary 2-D numerical tests show that it is best approximated with $\gamma=1.6$ for strong ductile target materials (Davison et al., 2011), while for target materials with no or little cohesion, it is not the case (Elbeshausen et al., 2009).

10.10.4.2.4 Impact heating, melting, and vaporization

The essential part of the initial projectile kinetic energy transforms into target material thermal energy, distributed around the crater. The heating is the result of energy conversion of the residual internal energy after the shock compression and decompression cycle and depends on the shock-wave amplitude. The further from the point of impact, the smaller impact heating.

Close to the point of impact, the zone of the originating impact crater cavity, heat large enough to change the thermodynamic state of the target material above its boiling point is deposited by the irreversible (unisentropic) process of shock-wave compression. However, for typical impact velocity of 20 km s^{-1} and for typical terrestrial planet crustal rocks, such as granite, basalt, and anorthosite, impact heating is not strong enough to fully vaporize the rocks. Most probably, a state of partial vaporization is reached, where rock melt at liquidus temperature (pressure-dependent) is in equilibrium with gaseous phases. The detailed thermodynamic history for different target materials can be calculated using *equation of states*. During partial vapor cooling, the chemical content of liquid and vapor phases can vary with a possible quenching at about 2000°K. In experiments, a difference in condensation products is observed under cooling and expansion after the shock-induced nonequilibrium process, when compared to quasi-equilibrium processes. Laboratory data about simulated rock vaporization are illustrated by Gerasimov et al. (1999). The presence of various oxides with temperature-dependent chemistry, for example, can be used to constrain the models (e.g., Yakovlev and Basilevsky, 1994).

For residual temperatures below the boiling point, shock heating results in *impact melting* of rocks. In terrestrial craters, impact rocks are easily identified by texture. The important differences of impact melt in comparison with volcanic melts are (1) the presence of large volumes of material melted above the solidus temperature (overheated melt) and (2) the presence of heavy fragment loading, untypical for volcanic melts. The history of fragments (clasts) incorporated in the impact melt may be very complicated – fine dust may be digested by melt, while large boulders seem to hover inside the quickly solidified melt body. The classification of impact-related metamorphism and rock types is still not fully settled, but some schemes and glossaries exist (e.g., Stöffler and Grieve, 2007).

With the exception of very large craters with diameters of about 200 km and larger, impact melt solidification has a two-stage cooling history (Onorato et al., 1978). During the first stage of cooling, the local thermal equilibrium with incorporated clasts is achieved. Later, the equilibration (cooling) of the entire melt body occurs by convection (in large craters) and heat conduction (in smaller craters), where local thermal equilibrium results in

temperatures close to solidus and correspondingly large viscosity. In terrestrial craters with diameters below 100 km, such as Manicouagan, clast cooling decreases the impact melt temperature to solidus in about 100 s. On Venus, the local thermal equilibrium is reached slower due to the smaller thermal gradient between clasts and melt and increases the final melt temperature, because Venus has a larger crust temperature (~740 K) than Earth. Hence, on Venus, impact melt is mobile an order of magnitude longer time (Ivanov et al., 1992). In large impact craters on Earth, such as Sudbury, the volume of impact melt is so large that thermal equilibrium is already reached at temperatures above the solidus and even above the liquidus, so that magmatic processes in the melt body, including convection, can continue (Zieg and Marsh, 2005).

Experimentally, the vaporization or melt volume has not been determined, because velocities under laboratory conditions are not yet high enough. Instead, impact melt volumes of a few terrestrial impact craters have been estimated (Cintala and Grieve, 1998a,b; Grieve and Cintala, 1992) and were compared with theoretical values obtained from shock-wave propagation and crater scaling laws. Today, this is the only method available to verify scaling laws for large impact craters. Typically, the melt volume or mass is given relative to the projectile volume or mass. A comparison made by several authors gives the estimated accuracy of a factor of 2 for the projectile mass and a factor of $2^{1/3}$ for the projectile size, respectively, for assumed projectile density, impact angle, and velocity (Ivanov, 1981; Pierazzo et al., 1997; Ugalde et al., 2005). **Figure 5** illustrates the comparison between impact melt volumes observed and estimates from scaling laws for terrestrial impact craters in a wide range of crater sizes. In the presence of porous target rocks (e.g., sandstone), the melt volume may be larger than in a nonporous rock (Wünnemann et al., 2008).

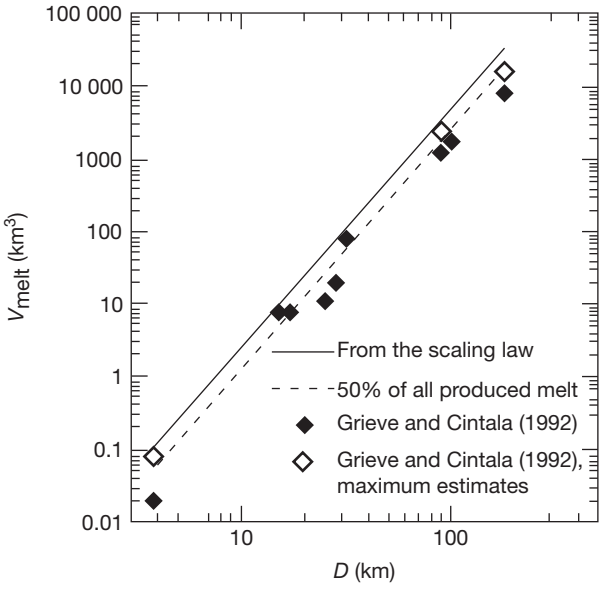


Figure 5 Comparison of estimates made using scaling laws (solid and dashed lines, constructed using data by Ivanov (1981) and Pierazzo et al. (1997)) and using geologic mapping of terrestrial impact craters, summarized by Grieve and Cintala (1992).

In summary, scaling laws are valuable rules to describe the morphometry and volume properties of impact structures with respect to the projectile. However, many details are still unexplored and often rely on either experimental or numerical approaches, and only the very small crater diameter range is covered by both. For the gravity-dominated regime, scaling laws can be highly unreliable, but it is impossible to reproduce natural impact craters by experiment. On the other hand, for small craters, the effects of the projectile energy or momentum coupling to the target, porosity, and effective strength are measurable parameters, which need to be implemented in numerical codes correctly. Although there are these numbers of uncertainties, the crater scaling is useful to study cratering distributions on the surfaces of the different planetary bodies.

10.10.5 Cratering Statistics

The bombardment history of the Moon is closely related to our understanding of planetary formation and evolution of the solar system. It turned out that the Moon is an ideal test site to study the cratering records in space and time, not only because most lunar endogenic activity ended more than 3 Ga ago, gradually decaying in time (e.g., Hiesinger et al., 2000, 2003). Over the last 3 billions of years, impact cratering alone dominated the modifications of the lunar landscape. Not only the Moon but also Mercury, Mars, and many of the giant planets' satellites kept the bombardment history of the earliest evolutionary period preserved on their surfaces, which were formed shortly after the formation of the solar system, but only the Moon has been studied by space missions extensively (space mission orbital dynamics and instrumentation are discussed in **Chapters 10.22** and **10.23**, respectively).

The Moon is the only other solar system body for which samples of known sampling site have been studied in detail, such as mineralogy and isotope ages. Absolute timescales for planetary surface evolution in the inner solar system (except the Earth) can only be derived by linking the lunar cratering frequencies with isotopically dated lunar samples. The approach of dating planetary surfaces by crater densities (i.e., the number of craters per square kilometer) was first proposed by Öpik (1960), whereas linking crater densities with isotopic ages was pioneered by Shoemaker et al. (1962). The Apollo and Luna sampling missions provided lunar rock samples that were isotopically dated; these isotopic ages were later correlated to crater densities on the lunar surface (Baldwin, 1971; Bloch et al., 1971; Hartmann, 1970a,b; Neukum et al., 1972, 1975b; Soderblom and Boyce, 1972). This permitted dating of surface features so that the planetary evolution could be studied through geologic time. These lunar cratering chronology models have not been updated since the 1970s. The most recent orbital evolution models (Bottke et al., 2012; Morbidelli et al., 2012b) compare these cratering-rate estimates for the Moon (i.e., Neukum et al., 1975a) to calibrate modeled flux evolution. However, one of the major criticisms is that these old cratering-rate estimates predict contradicting sequences of basin formation ages when compared to isotope dates.

The discovery of extrasolar planets has changed ideas on the dynamical evolution of our own planetary system and has inspired to develop the orbital migration of the gaseous planets during their primordial accretion phases (originally suggested by Zharkov and Kozenko, 1990) further. Through inward and outward migration during their formation period, these giant planets destabilize the remaining planetesimals (Gomes et al., 2005; Tsiganis et al., 2005). This could explain not only (1) what was earlier called LHB of the terrestrial planets, a spike in the inner-planet bombardment rate at around 3.9 billion years ago, originally postulated for the Moon from the isotope geochemistry of returned lunar samples (Tera et al., 1974), but also (2) the unusually small size of Mars (Hansen, 2009). Commonly used cratering statistics for planetary surface-age determination assume monotonic cratering-rate decay, but this assumption has been suggested to be dynamically unjustified. Hereby, more criticism is raised against the standard crater-counting method to determine planetary surface ages.

The possible nonsmooth cratering rate consisting of 'spikes' and 'spike tail' periods is the natural assumption, and the most probable graphic expression of the cratering rate versus time is a set of 'saw teeth' (Hartmann, 1970b). Recent modeling reveals a possible timescale of a 'spike' decay after a prominent injection of asteroids into planetary orbits of the order of 150–200 My (Bottke et al., 2007) and provided a new family of asteroids, which appeared in orbits near 'transporting' resonances. For an individual crater-retention age, the assumption of a constant (or monotonically decaying) impact rate could result in an erroneous absolute crater-retention age. However, the general stratigraphic approach (the more primary craters, the older the surface) remains a good assumption in planetology.

The crater-counting method, however, is the only tool to date planetary surface evolution when samples are not available. Uncertainties in the cratering rates and the diverse crater-production function shapes used to calibrate crater density measurements, as well as the assumed flux evolution, have resulted in many unique lunar chronology models. This diversity of lunar chronologies propagates to other planets such as Mars and adds ambiguity to the interpretation of their geologic history (e.g., Werner and Tanaka, 2011). Here, we describe some of the standard characteristics of cratering chronology model assumptions and crater SFDs, and we highlight the issues and challenges of the determination of the ultimate crater-production function (a mathematical expression for the original projectile distribution).

10.10.5.1 SFD of Impact Craters

Prerequisite for the determination of the lunar cratering chronology is the determination of the original crater SFD. There are a number of well-known and long-standing difficulties for the determination of the crater-production function. These difficulties not only are partly inherent issues of the cratering process but also and more significantly are related to the interpretation of the crater record. Numerous crater-production functions exist, because the major challenge in the development of the crater-counting technique and for surface-age determination is to find large, homogeneous, and undisturbed surfaces and to determine the true crater-production function. Observed crater distributions could be diverted due to the following:

1. *Average impact velocities*, which compress or stretch the curve along the diameter axis
2. *Resurfacing processes*, which shallow the curve
3. *Admixture of secondary craters*, which steepens the curve
4. *Target properties*, which modify the curve at the small crater range
5. *Mapping and counting* (average and distortion), when geologic boundaries are ignored

The implications will be discussed in further detail at the end of this chapter.

The statistics of cratering on planetary surfaces are commonly described by the shape of the SFD (e.g., **Figure 6**). The theoretical concept and mathematical background for dating techniques based on crater counts were developed in the 1960s and 1970s, as summarized by the *Crater Analysis Techniques Working Group* (Arvidson et al., 1979). Histogram graphs, used to display crater statistics, plot *cumulative* crater frequencies (number of craters greater than or equal to a given size per square kilometer, N_{cum}) against the crater diameter D on log–log diagrams. Distribution characteristics are defined by polynomial functions or multiple-segment power laws, where the cumulative frequency N_{cum} is proportional to D^{-b} and b varies between 4 and 1.5. There are other representations such as *incremental*, *differential*, or *relative* (R)-plots discussed in more detail in the succeeding text.

The main observation is that small craters are significantly more numerous than large craters, reflecting the proportion of the fewer large projectiles compared with the number of small projectiles in the solar system (e.g., Baldwin, 1964), and that

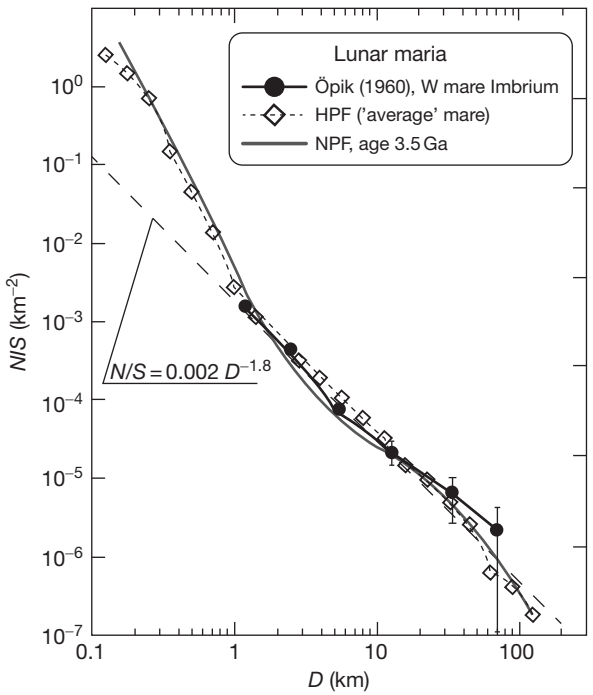


Figure 6 The early cumulative crater count (Öpik, 1960) on the lunar nearside in comparison with more extensive data published later.

has been recognized well before active Moon space-based exploration. Figure 6 illustrates the early crater counts on lunar maria using telescopic images of the lunar nearside (Öpik, 1960). Öpik's early data set shows that the cumulative number of craters in the diameter range from 1 to 100 km is close to a $D^{-1.8}$-dependence.

The SFD of craters or projectiles is the relation between the object size (e.g., diameter) and the number of such objects of the population investigated. There are three main styles of the SFD presentation (reviewed by the Crater Analysis Techniques Working Group (Arvidson et al., 1979):

1. *Cumulative* form of the SFD plot represent the number of craters larger than or equal to a given diameter, $N(\geq D)$.
2. *Differential* forms of the SFD plot represent the incremental number of craters in a given diameter bin (from the *left* bin-limiting diameter to the *right* bin-limiting diameter), $D_L \leq D < D_R$, as $\Delta N/(D_R - D_L)$.
3. *Logarithmic* forms of the differential SFD plot represent the incremental number of craters in a given logarithmic D-bin, where $D_R/D_L =$ constant and typically is $\sqrt{2}$. The slope (in near-linear segments) equals that of the cumulative SFD.
4. *Relative* form of the SFD plot represent the differential form, multiplied by D^3:

$$R = D_{av}^3(\Delta N/S)(D_R - D_L) \quad [19]$$

where the geometric mean interval diameter is defined as $D_{av} = (D_R \cdot D_L)^{0.5}$. The incremental number of craters is divided by the size of the counting area so that the R-value becomes nondimensional. However, for statistics of other object types, such as asteroids, the R-value can be a dimensional one. The graphic representation of the $R(D_{av})$ relations is known as the *R-plot*.

The different types of plots have different advantages and disadvantages and can be used complementarily: For example, cumulative SFD forms tend to smooth irregularities, such as deficiencies of craters in certain D-bins, and therefore, it is more difficult to detect crater obliteration processes. Because of the cumulative number, the statistical errors may become very small. Differential forms, essentially incremental histograms, have the advantage of showing the actual $N(D)$ behavior for each D-bin, including turndowns caused by crater obliteration. Conclusions derived from these plots, though, could be statistically less significant due to small-number statistics and large variance. The log differential forms have the advantage of mimicking the cumulative $N(D)$ slope, so that slopes of the distributions can easily be compared.

The ideal case to study the crater SFD is on a planetary surface that instantaneously begins to accumulate new craters and never gets modified. The observed crater SFD reflects the production SFD or *the standard distribution* of the projectiles. In natural cases, crater degradation and other obliteration processes change the population of these craters. To derive a production function from natural examples of crater SFD measurements, many authors have tabulated and generalized lunar crater counts. Early results are presented, for example, by Hartmann (1969), Hartmann (1970a), Neukum et al. (1975a,b), and Neukum (1983) and summarized in an early review paper by Hartmann et al. (1981), which developed into the two most commonly used varieties of the crater-based age-determination method. Therefore, we concentrate in this chapter on these two lunar *production* crater SFDs as they were proposed by Hartmann and Neukum (see details in Neukum et al. (2001a)) and then updated by Hartmann (2005). They are well described and easy to implement. The term *production function* of a specific age, or the *isochron*, has been defined as the SFD of craters forming on planetary surfaces for the period the planetary surface was exposed, which defines the age of the surface (e.g., Neukum and Ivanov, 1994; Strom et al., 1992).

10.10.5.2 Hartmann Production Function

In his work on crater SFDs, Hartmann focused on the Moon and Mars. He describes the observed crater distributions in a log-incremental representation with a standard $\sqrt{2}$-diameter bin size. The number of craters per square kilometers, here, is calculated for craters in the diameter bin $D_l \leq D < D_R$, where D_L and D_R are the left and right bin boundary and the standard bin width is $D_R/D_L = 2^{1/2}$. For the Moon, Hartmann defined power-law descriptions for three diameter intervals (Hartmann, 1999):

$$\log N_H = -2.198 - 2.20 \log D_L \quad \text{for } D_L > 64\,\text{km} \quad [20a]$$

$$\log N_H = -2.920 - 1.80 \log D_L \quad \text{for } 1.41\,\text{km} < D_L < 64\,\text{km} \quad [20b]$$

$$\log N_H = -2.616 - 3.82 \log D_L \quad \text{for } 0.3\,\text{km} < D_L < 1.41\,\text{km} \quad [20c]$$

Hartmann used for the larger craters ($D > 4$ km) an exhaustive catalog of individual crater-by-crater diameters, measured by Arthur et al. (1963, 1965a,b, 1966), documented in Figure 4 of Hartmann (2005), to determine an average mare production distribution. The condition to have a uniformly fresh surface is nearly satisfied by the fact that most lunar mare basalt samples have a narrow range of ages (i.e., 3.2–3.6 Ga; Hiesinger et al., 2003; Stöffler and Ryder, 2001), although the flux of crater-forming projectiles during that period dropped by a factor of 2–3. The observed crater distribution of the mare units reaches a saturation equilibrium areal density above crater diameters of $D \sim 300$ m (Hartmann, 1984). The assigned age for the average mare distribution was in later publications modified from 3.4 to 3.5 Ga (Hartmann, 2005; Hartmann and Neukum, 2001).

Deriving Mars isochrons, Hartmann refined his approach over some years by a set of iterations (1999, 2002, and 2004) (Hartmann, 2005; Hartmann and Neukum, 2001). He started by extrapolating the lunar slope for the diameter range of craters smaller than 300 m according to eqn [20c] (iteration 1999) and subsequently by adopting a continuation of the curve by grafting on the shape at the small-diameter end of the crater-production function as described with the polynomial expression by Neukum (e.g., Neukum and Ivanov, 1994). The slope representing the polynomial expression in that diameter range is the −2.9 cumulative slope (Shoemaker et al., 1970):

$$\log N_H = -2.0 - 2.90 \log D_L \quad \text{for } 0.01\,\text{km} < D_L < 0.125\,\text{km} \quad [20d]$$

Subsequently, Hartmann (2005) adjusted the iteration 2002 shape by the effects of atmospheric shielding for small Martian impact craters as it is calculated by Popova et al.

(2003). We call this result the *Hartmann production function*. The calibration of the absolute ages, and the calculation of the isochrons as examples are given in **Figure 7**, will be discussed in the succeeding text.

10.10.5.3 Neukum Production Function

A number of studies of the lunar crater SFD revealed that there is no unique power law to present the crater SFD in the whole range of crater diameters (Chapman and Haefner, 1967). Instead of using power-law segments such as Hartmann, Neukum et al. (1975a) proposed an analytic function to describe the cumulative number of lunar craters, N, per square kilometer with diameters larger than a given value D. They computed a polynomial of 7th degree by a least-square fit for a crater diameter range between 300 m and 20 km. In a series of publications – (see summary by Neukum and Ivanov, 1994) – this polynomial was extended to the 11th degree and for a diameter range between 10 m and 300 km. They proposed this polynomial to be applicable, considering crater scaling, on all terrestrial planets, asteroids, and some moons, and it is expressed by

$$\log_{10}(N) = a_0 + \sum_{n=1}^{11} a_n \log_{10}(\mathrm{D}^n) \qquad [21]$$

with $0.1\ \mathrm{km} < D < 300\ \mathrm{km}$ validity range for the Moon, where D is in kilometers, N is the number of craters with diameters $\geq D$ per square kilometers per billion of years, and the coefficients a_n are given in **Table 1** for the Moon. Recently, the coefficients were slightly reworked for the largest crater range by careful re-measuring in the size range (Neukum et al., 2001a). We call this result the *Neukum production function*. The time dependence of the a_0 coefficient is discussed in the following subsection.

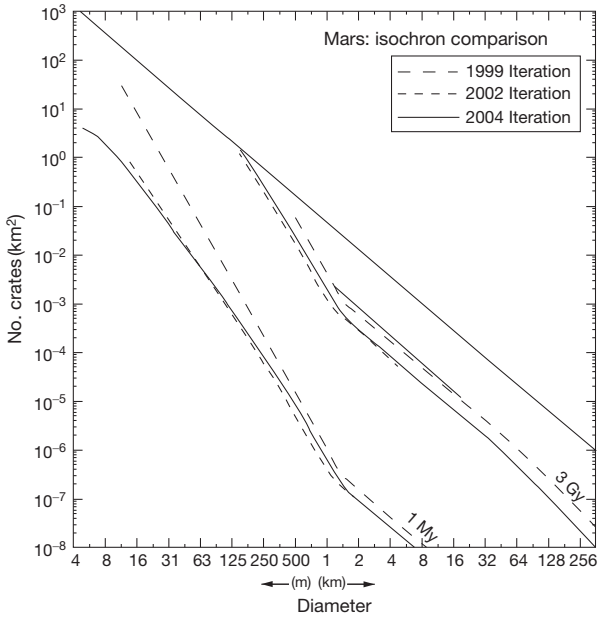


Figure 7 Iteratively evolved isochrons for 1 Ma and 3 Ga of the Hartmann scheme in comparison. Source: http://www.psi.edu/epo/isochrons/chron04b.html.

Although the shapes of both types of crater-production functions differ for crater diameters larger than 1 km and vary significantly for crater diameters between 2 and 25 km (e.g., Ivanov, 2001; Werner and Tanaka, 2011), they show a rather similar overall behavior (**Figure 6**). Despite the individual approach of these models, both assume that the general shape of the crater-production function for objects striking the Moon stayed the same for at least the past 4 Gy. This point of view has been challenged over the years, mostly by Strom (Strom and Neukum, 1988; Strom et al., 1992, 2005), who suggests that the modern (post-mare formation) production function is different from that produced during the epoch of the LHB (before about 3.9 Ga ago). Similarly, modern numerical models for the dynamical evolution of small bodies in the solar system support the possibility that the velocity and SFD of the projectiles may vary throughout the first billion years of solar system evolution (e.g., Bottke et al., 2012; Morbidelli et al., 2012b).

Several issues can occur in case of long-term exposure to cratering of a planetary surface. From the very beginning of the impact crater population studies, it was recognized that ongoing bombardment of a given surface with low little endogenic activity (like on the Moon) impact craters would cover the entire surface, so that any new crater formed would destroy older craters. Hence, the areal crater density (number of craters of a given size per unit area) stops to grow, reaching some maximum value defining the state of the so-called saturation equilibrium (various authors use this term or prefer to say briefly 'saturation state' or 'equilibrium state'). After the pioneering physical modeling by Gault (1970), many publications have clarified this problem, and are reviewed in the introductory part of the paper by Richardson (2009). Briefly, the study revealed two main mechanisms that limit the number of craters to be counted at or near the saturation limit. If the production function is *steep* ($N \sim D^{-n}$, $n > 2$), small craters tend to erase larger craters ('sandblast regime'). The steep production function is typical for lunar craters smaller than ~1.5 km. The sandblast regime results in the limiting slope of the crater SFD of $N \sim D^{-2}$ for all slope indices $n > 2$.

Table 1 Coefficients for eqn [21] for the lunar case

a_n	N_{cum} (D) (Neukum, 1983)	N_{cum} (D) (Neukum et al., 2001a)	Sensitivity (%)[a]
a_0	−3.0768	−3.0876	
a_1	−3.6269	−3.557528	±3.8
a_2	+0.4366	+0.781027	±3.9
a_3	+0.7935	+1.021521	±1.6
a_4	+0.0865	−0.156012	±1.6
a_5	−0.2649	−0.444058	±0.88
a_6	−0.0664	+0.019977	±1.3
a_7	+0.0379	−0.086850	±0.78
a_8	+0.0106	−0.005874	±1.8
a_9	−0.0022	−0.006809	±1.8
a_{10}	-5.18×10^{-4}	$+8.25 \times 10^{-4}$	±5.6
a_{11}	$+3.97 \times 10^{-5}$	$+5.54 \times 10^{-5}$	±24.1

[a]Sensitivity is the coefficient variation, which changes the N_{cum} (D) value a factor of 2 up and down.

For *shallow* production functions ($N \sim D^{-n}$, $n \leq 2$), the main mechanism is the episodic renovation of the area by the largest crater (the so-called cookie-cutting regime). After a regionally critical impact event, the adjacent area is renovated by the ejecta blanket and the fresh surface starts to accumulate craters again. The SFD of craters near equilibrium in the cookie-cutting regime reproduces the primary production function shape (Chapman and McKinnon, 1986; Woronow, 1978). This situation is observed in heavily cratered lunar highlands. The modern modeling e.g., Richardson, 2009 included more sophisticated models of crater evolution (like downslope motion from the crater wall) confirms that the shallow part of the primary production function should be still visible at heavily cratered terrains. However, these modern models do not include a possible change in the target properties as the upper layers are progressively affected by multiple impacts. More refined models should be constructed to take into account higher target temperature of the Moon during the earliest bombardment period (this may affect the largest craters; e.g., Ivanov et al., 2010b) and the subsequent decay of the bombardment rate (e.g., no basins younger ~3.7 Ga).

One of the important features of impact craters in saturation is their less-random distribution at the surface (Squyres et al., 1997). Primary impacts usually occur in random spots at the surface. The statistics of center-to-center distances may be mathematically simulated with by a random process. Craters in saturation (meaning: recognizable craters) reveal center-to-center distance statistics shifted to a more regular spatial distribution. Such features are not studied in the modeling approaches mentioned in the preceding text. Only few papers take this statistics into account (e.g., Michael et al., 2012).

A rule of thumb was proposed by Hartmann (1984) to draw a single power-law SFD for craters in the saturation/equilibrium with a relation of $N \sim D^{-1.83}$, which can be used for a primary visual inspection of the closeness to the equilibrium. In the notation explained in **Section 10.10.5.2**, Hartmann's empirical saturation line is expressed as $\log N_H = -1.83 \log D\,(\mathrm{km}) - 1.33$.

10.10.6 Impact Rate Estimates

Impact rate estimates for the earliest evolution of solar system planetary bodies can be obtained only from models, and thus, they depend on the theoretical concept of solar system evolution, assuming that the bombardment history of the Moon is closely related to our understanding of planetary formation and evolution of the solar system. For the period about 4.1 Ga and younger, rates can be determined by correlating crater frequencies measured near Apollo and Luna sampling sites and absolute apparent ages for the returned samples. Among others (Baldwin, 1971; Bloch et al., 1971; Hartmann, 1970a; Neukum et al., 1972; Soderblom and Boyce, 1972), both Hartmann (1970a) and Neukum et al. (1975a) had shown that crater densities at lunar landing sites of different ages require that the cratering rate decreased by a factor of 100 between 3.9 and 3.3 Ga ago, implying a cratering-rate half-life of the order of 10^8 years. This permitted dating of surface features so that the planetary evolution could be studied through geologic time.

Modern impact rates on planets, considered for the period from about 3 Ga ago to today, are assumed to be a result of a quasi steady-state regime. Rates are defined by orbital evolution of small bodies under weak gravitational forces including resonant perturbations of the large planets acting dominantly on asteroids in the MAB and close encounters on planets in the inner solar system, such as the near-Earth asteroids. Additionally, nongravitational (nonconservative) forces act on smaller asteroids (diameter <20 km). Such are the effects of solar radiation, which produces a radial migration if the asteroid rotates. However, collisions between asteroids, or small bodies in general, have been earlier evoked as a process that not only shaped the current and past asteroid SFD down to particles of dust size (which can be removed from the solar system by radiation pressure of the sunlight) but also could be a mechanism to change orbital parameters of fragments to replenish the *Kirkwood gaps* in the MAB orbital parameter space so that a fraction of fragments reach planet-crossing orbits, producing small meteoroids, which could be found on Earth as meteorites and crater-forming projectiles. (Kirkwood gaps are zones in the orbital parameter space, which are not populated, because objects with such a parameter range are perturbed mostly by gravitational modification due to orbital resonances with not only Jupiter but also Saturn; for more details, see **Section 10.10.5.1**.)

Collisions affect the physical state of asteroids, orbital parameters, and intrinsic angular momentum, forcing fragments into orbits that are easily affected by resonant perturbations, so that the fragments end up in planet-crossing orbits. The extreme case is a catastrophic collision, resulting in the asteroid disruption. The largest fragment is produced with a mass of one-half of the initial asteroid mass (Greenberg and Hartmann, 1977; Greenberg et al., 1978). For small asteroids (with diameters smaller than about 1 km), the resulting fragment-sized distribution is controlled mostly by the strength of the asteroid material. The effective strength is usually expressed in units of the specific energy of the impactor per unit mass of the target body. For laboratory-scale rocky model bodies, the catastrophic specific energy is in the range of 300–2000 J kg^{-1} (O'Brien and Greenberg, 2005, see summary in their Figure 4a). With increasing asteroid size, the strength of an asteroid declines slowly, because larger bodies in nature commonly contain more intrinsic defects (fractures, pores, and inhomogeneities), and thus, the resistance to deformations and fragmentation is decreased (Love and Ahrens, 1996; Ryan and Melosh, 1998). For large asteroids (with diameters of 1 km and larger), self-gravitation becomes important, keeping fragments of an impact event together, and the dispersal of fragments is inhibited. Such an accumulation of fragments, kept together due to gravity, is nicknamed rubble pile, and they are asteroids, repeatedly shattered but not dispersed. The asteroids in the size range of 200–800 m have the smallest strength against disruption (O'Brien and Greenberg, 2005, see summary in their Figure 4b).

The frequency of giant impacts in the MAB constrains the number of large crater-forming projectiles. The rarity of impacts that produce fragments larger than 1 km and divert them toward planets, as correctly pointed by Shoemaker, resulted in an early assumption that most of impact craters with diameters larger than ~30 km on terrestrial planets are

formed by comets (Shoemaker and Wolfe, 1982; Shoemaker et al., 1990). The recognition of a prevailing role of gravitational forces (resonances) in the repopulation of planet-crossing orbits from the MAB (Farinella et al., 1994) has destroyed Shoemaker's main argument for comet dominance. Modern estimate of the comet fraction in the impact cratering is about $10 \pm 5\%$ (Bottke et al., 2002, 2005; Weissman et al., 2002). Delivering frequencies derived with the Nice concept of planetary system evolution, comets play a negligible role (Bottke et al., 2012).

The lifetime of bodies at planet-crossing orbits was assumed to be controlled by close encounters with planets (Arnold, 1965; Wetherill, 1967). From the modern point of view, this model gives too high estimates of the lifetime of objects on planet-crossing orbits (Dones et al., 1999). Gravitational resonances are now seen to be a major factor in producing objects with planet-crossing orbits.

10.10.6.1 Gravitational Mechanisms

One critical mechanism in changing small-body orbits into planet-crossing orbits is the resonant perturbation of giant planets. There are two broad kinds of resonances (Morbidelli, 2002): (1) *mean motion resonances* (the periods of rotation around the Sun of two bodies give the ratio of integer numbers) and (2) *secular resonances* (the precession frequency of the asteroid's longitude of perihelion is equal to the integer number of a secular frequency of the planetary system). Especially, resonances with Jupiter create gaps, *Kirkwood gaps*, in the MAB orbit distribution as a function of semimajor axis a. The most prominent mean motion resonances with Jupiter form gaps at mean orbital ratio of 3:1 ($a = 2.5$ AU), 5:2 ($a = 2.82$ AU), and 2:1 ($a = 3.27$ AU) between the asteroid and Jupiter. For the deflection of main belt asteroids (MBA) into planet-crossing orbits, two resonances are found to be the most important, the 3:1 mean motion resonance with Jupiter and the secular ν_6 resonance with Saturn. (Michel et al., 2005). Many other resonances influence asteroid orbit evolution (Milani and Farinella, 1995; Morbidelli and Gladman, 1998).The action of resonances results in a quick increase or chaotic change of an asteroid's eccentricity, making it a planetary orbit crosser. While most of the resonant perturbations destabilize asteroid orbits, the *Kozai resonance* (which occurs when the precession of the asteroid's longitude of perihelion equals the precession of its node and keeps the quantity $a(1-e^2)^{0.5} \cos i$ constant during a planetary encounter) provides a sort of protection from close encounters. Bodies with low inclination and a semimajor axis close to that of Earth or Venus may enter the Kozai resonance, which causes the perihelion to librate around 0° or 180°. Moreover, the semimajor axis remains almost constant, which protects the body from a collision with the planet (Gladman et al., 1995; Michel and Thomas, 1996; Michel et al., 1996; Milani et al., 1989), at least temporarily.

10.10.6.2 Nongravitational Mechanisms

The orbits (1) and the rotation rates and obliquity (2) of the small bodies in the solar system may be modified by the thermal effects and the radiation forces. Both parameters are influenced by solar heat, accumulated by the day side of an asteroid and reradiated from the surface when a given area rotates to the night side (Bottke et al., 2006; Morbidelli and Vokrouhlicky, 2003; Rubincam, 1998). The effect depends on thermophysics of the thin surface layer, distance from the Sun, and obliquity of the asteroid. The modification of the orbit of a small body is called *Yarkovsky effect* (1) and acts as not only a function of size and heliocentric distance but also material properties. For a typical 1 km body, the Yarkovsky effect results in the orbital semimajor axis drift of the order of 10^{-4} AU per 10^6 years. The comparison of modeled Yarkovsky effect with astronomical observation is still limited to a few asteroid observations (Chesley et al., 2003) and statistical investigations of asteroid family evolution (Vokrouhlicky et al., 2006a,b). The Yarkovsky effect operates slowly, but it is important for the MBA evolution and delivery of small bodies to resonant orbits.

The solar-induced thermal torque may also change the rotation rate and spin axis orientation of the small body. The related *Yarkovsky–O'Keefe–Radzievskii–Paddack* (YORP) *effect* (2) explains how the solar irradiation may change the spin state (spin up or spin down) of a small body by the reflection and reradiation of absorbed sunlight (Capek and Vokrouhlicky, 2004; Rubincam, 2000). The strength of the effect depends on the size of the object and its obliquity; however, the full implications of this effect require further studies.

Both effects require too long timescales, to be considered as important delivery mechanism outside the asteroid belt. Once an object becomes a planet-crosser, the main disturbing events are encounters with terrestrial planets, and the Yarkovsky effect has no time to build up the necessary input for the orbit evolution (Michel and Yoshikawa, 2005).

While modeling the dynamical evolution from the MBA to NEO, nongravitational effects predict a steepening of NEO SFD versus the MBA SFD due to increased small asteroid *mobility* in the phase space, resulting from the Yarkovsky and YORP effects (Morbidelli and Vokrouhlicky, 2003). These effects are believed to dominate in decameter-sized range (e.g., Jedicke et al., 2002, their Figure 7) but affect different taxonomic types differently (Hartmann et al., 1999). Nongravitational effects on NEO SFDs are yet visible only as a tendency for absolute magnitudes $H < 18$. However, radiation-driven models offer progress in the understanding of asteroid family evolution (Bottke et al., 2006; Vokrouhlicky et al., 2006a,b). Being one of the fundamental causes for asteroid orbital evolution, which, thus, are important for the impact-rate discussion, the Yarkovsky and YORP effects still estimate the absolute rate of evolution with only limited accuracy, because the thermophysical properties of asteroids are poorly known. For example, the age of the Baptistina family was originally estimated in the range of 140–190 My (Bottke et al., 2007) but was later adjusted to the 140–320 My range, due to possible density ($\pm 30\%$) and thermal conductivity (factor of 3 up and down) variations (Masiero et al., 2012).

10.10.6.3 Impact Probability

The impact probability of NEOs is a parameter that connects the observed NEO population and the number of actual impact events on the target planet, measured as the recent cratering rate (e.g., Oberst et al., 2012) or as the rate of bolides recorded as light flashes (meteors or bolides or fireballs) in the

terrestrial atmosphere (e.g., Ivanov, 2006). Key assumption here is that the population of NEOs, N_{NEO}, is in a steady state. The average (for the whole NEO population) probability of collision with the target planet is P_{col} (measured in year^{-1} per one object); the impact rate, R_{col}, defined as number of impacts (collisions) in the specific time period (e.g., per year) is

$$R_{col}(\text{year}^{-1}) = N_{NEO} \cdot P_{col} \quad [22]$$

Standard approaches estimate encounter probabilities, P_{col}, according to Öpik's formulas (Bottke and Greenberg, 1993; Öpik, 1951, 1976; Wetherill, 1967) provided that the main orbital parameters for NEOs and the target planet, a, e, i, are constant in time and all angles defining mutual position of orbital planes are equally probable. Typical encounter probabilities, P_{col}, are in the range of 10^{-11}–10^{-8} year^{-1}, and during a typical lifetime of the individual NEO of 10^7 year^{-1}, the probability of impact is very low. In reality, individual objects change their orbital parameters across a wide range of the phase space during their lifetime. Hence, eqn [22] should be treated as an estimate for only a specifically populated orbit set that requires the population to be in a steady state. This state is reached if, for example, (1) an Apollo object changes its orbit to the Amor-type orbit and another NEO changes its orbit from the Amor-type to Apollo-type orbit (see Milani et al., 1989) and (2) the number of objects leaving the NEO family, for example, striking the Sun and planets or being ejected out of the inner solar system, is replenished with new NEOs arriving from the MAB orbits.

Estimates using forward integration of known objects considering perturbations as discussed in **Sections 10.10.5.1** and **10.10.5.2** also yield impact probabilities. The forward integration of Earth-orbit-crossing asteroids with $H < 18$ produces statistics of close approaches (distance < 0.1 AU) with average impact probability of all NEOs (bodies with $q < 1.3$ AU) as low as 1.7×10^{-9} year^{-1} per one NEO, which is equivalent approximately to 3.4×10^{-9} year^{-1} per individual Earth-orbit-crosser (Harris, 2003; Morrison et al., 2002; Stuart and Binzel, 2004). Most of the astronomically observed NEOs are tested by numerical orbit integration to be estimated as a potentially hazardous object (PHO). The list of PHO close encounters with Earth is published by several groups. None of the tested NEOs, according to the numerical orbital motion integration, show a chance of striking the Earth or Moon in the historically short time, with the exception of 99942 Apophis. However, highly accurate impact probability estimates require good knowledge of orbital parameter and physical properties to calculate all perturbation effects correctly. Currently, an impact probability greater than 10^{-6} is found for an impact in the year 2068 (Farnocchia et al., submitted for publication).

Parallel estimates by numerical simulations and by Öpik-type estimates for a large population of asteroids based on cratering-rate derivations may be treated as the average probability of steadily populated planet-crossing orbits. One challenge of estimating the impact probability is that the subset of orbits having the largest collision probability simultaneously are orbits with the shortest lifetime. Therefore, the high probability of a close encounter can result in either collisions or an abrupt orbit change during the close encounter. Hence, there is only a small chance to observe NEOs in these orbits. By analyzing real collisions of small and hence frequent enough bodies colliding with the Earth atmosphere and observed as bolides (Morbidelli and Gladman, 1998), the statistical investigation of impact probability for small NEOs can be extended. Systematic photographic surveys of bolides have been made in different countries in different periods of time (Borovicka et al., 2005; Ceplecha, 1987; Halliday et al., 1996; McCrosky et al., 1978; Oberst et al., 1998). Registration of each event from several observational points allows researchers to estimate orbital parameters just before the atmospheric entry. Using these orbital parameters, one can make a formal calculation of the Öpik-type collision probabilities and compare these estimates with similar values for astronomically observed NEOs, by assuming an appropriate NEO SFD. The latter is one of the crucial reasons why these bolide impact probability data may be diverting from large astronomically observed objects. Only NEOs with diameter of ~1 km and larger have close to complete statistics, but the statistics of bolides for small bodies with the effective diameter of a few meters and smaller are nearly undetermined (see discussion by Morrison et al., 2002). The Chelyabinsk meteor (15 February 2013) unexpectedly entered the atmosphere and exploded, making it the most recent and prominent example of undiscovered small projectiles. In addition, there are many inherent observational biases for bolide observations as a statistical database (Halliday et al., 1996). To reduce the severity of possible biases, estimates for bolides with various minimum photometric mass (0.1, 1.0, and 10 kg) from the database for bolides from many surveys need to be considered (Lindblad et al., 2003). Recent long-term and high-resolution space-based image coverage of the Moon and Mars allows discovering newly formed impact sites, which allow adding more numbers to the impact probability at the small-sized range of planet-crossing objects (Daubar et al., 2013; Oberst et al., 2012). As for all these estimates, uncertainties determining the projectile size are given, due to atmospheric breakup, crater scaling laws, and general unknown observational bias, as well as the challenge to link the SFD of these meter- or less-sized objects and the astronomically observed kilometer-sized NEO population (e.g., Morrison et al., 2002).

To check the idea that the small number of meteoroid collision with Earth occurs from a small number of orbits with high impact probability, data for NEOs and bolides are sorted in the order of increasing P_{col} with calculation of the accumulated average $<P_{col}>$ for the meteoroid ensemble:

$$<P_{col}(n)> = \frac{\sum_{i=1}^{n} P_{col,i}}{n} \quad [23]$$

where i is the number of the orbit ($i = 1, N$) and $P_{col,i}$ and $<P_{col,i+1}$. In this construction, first, the lowest-probability orbits are summed and the average value grows as $P_{col,i}$ monotonically increases with the number i in this sorted list. If a low number of orbits with higher collision probability give disproportional large input, the function $<P_{col}(n)>$ would rise when n approaches the total number of recorded events, N. **Figure 8** presents the relation between $<P_{col}(n)>$ and $P_{col,n}$, the largest probability in the partial sum in eqn [23] for Earth-crosser objects in asteroid-like orbits ($T_J > 3$, calculated according

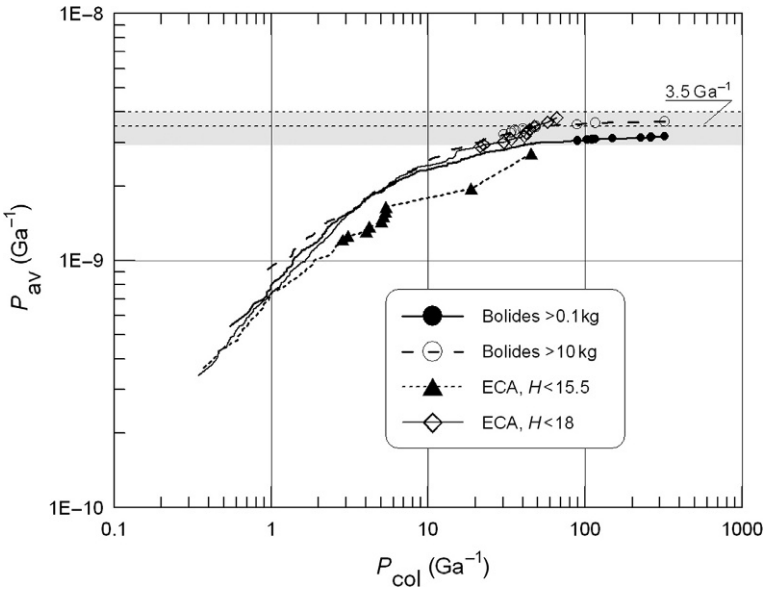


Figure 8 The accumulated sliding average for Earth-crossing asteroids (ECAs) and bolides, constructed by sorting the orbits according to their ascending collision probability (eqn [23]). All curves flatten for sparsely populated orbits with large collision probabilities. The average probability for the whole observed population is about 3.5 ± 0.5 Ga^{-1} per one Earth-crosser. For each data set, ten events with the largest collision probability are marked, to illustrate their importance.

Table 2 Estimates of collision probability and average impact (preatmospheric) velocity of bolides from meteor ground surveys

	Known Earth-crossers		*Observed bolides*		
	H < *18*[a]	H < *18*[b]	m_{ph} > *10 kg*[c]	m_{ph} > *1 kg*[c]	m_{ph} > *0.1 kg*[c]
$T_J > 3$					
N	401	461	74	152	227
P_{col}, year^{-1}	3.8×10^{-9}	3.52×10^{-9}	3.7×10^{-9}	3.2×10^{-9}	3.2×10^{-9}
$<U>$	19.0	19.0	19.0	19.8	20.3
$2 < T_J < 3$					
N (% of total)	54 (12%)	94 (17%)	18 (20%)	48 (24%)	106 (32%)
P_{col}, year^{-1}	0.80×10^{-9}	0.94×10^{-9}	0.88×10^{-9}	0.90×10^{-9}	0.93×10^{-9}
$<U>$	26.4	24.6	27.5	25.2	25.5
All objects with $T_J > 2$					
N	455	555	92	200	333
P_{col}, year^{-1}	3.4×10^{-9}	3.09×10^{-9}	2.3×10^{-9}	2.0×10^{-9}	1.8×10^{-9}
$<U>$	19.3	19.3	20.7	21.6	22.1

[a]Column as the MPC catalog (file 'astorb.dat') in November 2005.
[b]Column as the MPC catalog (file 'astorb.dat') in February 2013.
[c]IAU MDC Photographic Database, Version 2003, (Lindblad et al., 2003).

eqn [24] in the succeeding text) with *absolute magnitudes* (the absolute magnitude of asteroids, H, is defined as the apparent magnitude that the object would have if it was one astronomical unit (AU) apart from both the Sun and the Earth and at a phase angle of zero degrees) $H < 15.5$, for which completeness is assumed, and $H < 18$ (roughly >1 km in diameter, as of February 2013) objects and for fireballs with photometric masses more than 0.1 and 10 kg (as of 2003; Lindblad et al., 2003). One can see that the function $<P_{col}(n)>$ for $H < 15.5$ bodies and fireball meteoroids grows smoothly to the average collision probability per object of 3.5 ± 0.5 Ga^{-1}. These curves flatten for the largest probabilities, converging to the limiting average probability. Data for $H < 18$ objects may be corrected for the incompleteness of astronomical survey for these bodies. Possible undiscovered bodies with $H < 18$ will have orbits with small collision probabilities (low probability of close encounters) (compare Table 2), and therefore, the average collision probability will decrease, because a larger number of bodies with assumed low collision probabilities do not essentially add to the sum of individual orbital collision probabilities.

Celestial mechanics reveals two broad classes of orbits of small bodies in the solar system depending on whether the

orbit allows close encounters with Jupiter or not. Astronomically recent close encounters are reflected by the specific value of the Tisserand parameter

$$T_J = a_J/a + 2\cos I\left[\left(1-e^2\right)\left(a/a_J\right)\right]^{1/2} \quad [24]$$

where a, e, i are, correspondingly, semimajor axis, eccentricity, and inclination of the small-body orbit and a_J is the semimajor axis of Jupiter.

Table 2 compares Öpik-type collision probability estimates for orbits of observed bolides (as of 2003) and similar estimates for large Earth-crossers with $H<18$ (as of 2005 and 2013). Estimates show that for objects (meteoroids) in asteroid-like orbits ($T_J>3$), the average collision probability and average collision velocity for meteoroids, observed bolides entering the terrestrial atmosphere, are close to values for known Earth-crossers, with $P_{col}\sim(3\text{–}4)\times10^{-9}$ year^{-1} per body and $<U>\sim$19–20 km s^{-1}. This corresponds well with the results of orbital integrations and bias-corrected estimates (Morrison et al., 2002; Stuart and Binzel, 2004). Objects on comet-like orbits ($2<T_J<3$) appear to grow in proportion for smaller-sized objects, less than 20% for NEOs with $H<18$ to more than 30% for bolides with small photometric masses. An observational bias cannot be excluded and deserves still further investigation. The average collision probability, P_{col}, is significantly lower, and the average impact velocity, $<U>$, is slightly higher for objects on comet-like orbits, with P_{col} of about 0.9×10^{-9} year^{-1} per body, and the average velocity ranges between 25.5 and 27.5 km s^{-1}. The average values, especially for the collision probability, for all Earth-crossing asteroids and largest bolides are slightly different due to proportion of bodies striking Earth from asteroid-like and comet-like orbits (Table 2).

We can conclude that collision probability estimates based on fireball and observed asteroid databases are consistent and that observed and documented fireball data are in agreement with the celestial mechanics results. Similar estimates of collision probabilities for the other terrestrial planets are compared in Table 3. The collision probability and average impact velocity estimates are based on the known population of planetary crossers with $H<17$.

One can state that the bolide ratio within a factor of 2 is the same for Mercury, Venus, Earth, and the Moon. Mars, being close to the MAB, suffers three to four times more frequent impacts. However, the larger surface gravity and smaller average impact velocity on Mars in comparison with the Moon result in smaller crater sizes, created by projectiles of a given size. Hence, the ratio of the cratering rates and the frequency of the same crater size formation at the unit surface area on Mars and the Moon is smaller than the bolide ratio. Thus, the formation of an impact crater of a given size on Mars needs larger, consequently less numerous, projectiles than on the Moon (Ivanov, 2001).

10.10.7 Extraterrestrial Surface Dating

10.10.7.1 Principles of Surface Dating by Crater Counting

The extraterrestrial surface dating method, which is based on cratering statistics, has been developed by several groups since the early 1960s (Öpik, 1960). Dating of cratered planetary surfaces requires combining the information from geologic mapping and crater count statistics. Mapping delineates materials by geomorphic, multispectral, or reflectance features based on natural breaks in the rock sequence, while crater count statistics provide a quantitative tool for defining relative ages and absolute model ages. The theoretical concept and mathematical background for dating techniques based on crater counts were developed in the 1960s and 1970s, as summarized by the *Crater Analysis Techniques Working Group* (Arvidson et al., 1979). The determination of absolute model ages on the Moon (and other solid surface objects) is based on the application of a crater-production function (Section 10.10.4; several examples for the Moon are shown in Figure 9) and a cratering chronology model (Figure 10, for the Moon, e.g., by Stöffler and Ryder, 2001). The crater-production function describes the expected crater SFD observed on a geologic unit at a specific time and is, when scaled, similar for all planetary bodies (e.g., Hartmann, 1973, 1977; Hartmann and Neukum, 2001; Hartmann et al., 1981; Ivanov, 2001; Ivanov et al., 2001;

Table 3 Average impact probability and average impact velocity on terrestrial planets for observed planetary crossers $H<17$, listed in the MPC catalog in February 2013

Planetary body	$N_{H<17}$	*Impact rate*[a] *(averaged)* (10^{-15} *km*$^{-2}$ *year*$^{-1}$)	*Collision probability per one body (planetary crosser, averaged)* (Ga^{-1})	*Impact velocity (averaged)* (km s^{-1})	R_b[b]
Mercury	31	0.684	1.651	34.0	0.82
Venus	95	1.002	4.856	23.3	1.20
Earth	207	1.398	3.454	19.1	1.67
The Moon	207	0.835	0.153	17.2	1
Mars, $e=0.0934$ current orbit	3159	4.248	0.195	9.6	5.09[c]
Mars, $e=0.005$[d]	1684	2.717	0.234	10.2	3.25[c]
Mars, $e=0.001$	1013	2.282	0.326	10.5	2.73[c]

[a]Average impact rate is the global number of impacts per time unit, divided by the surface area of a planet.

[b]The bolide ratio is the ratio of the impact rate of projectiles of the same size per unit area to the same value for the Moon (for the Moon $R_b=1$ by definition).

[c]Not corrected for the incomplete observation of Mars-crossers (see Ivanov, 2001, for an example of the correction).

[d]This value of eccentricity is close to the average value for a time span of 10 Ma, still limited to the available Mars orbit modeling duration (Ward, 1992).

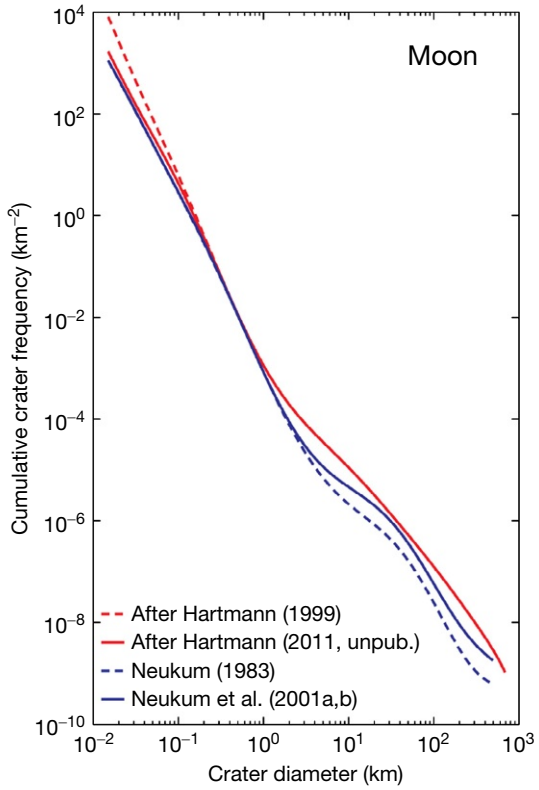


Figure 9 Four different cumulative crater size–frequency distribution (SFD) shapes in log–log scale for the Moon.

Neukum and Hiller, 1981; Neukum and Ivanov, 1994; Neukum and Wise, 1976; Soderblom et al., 1974; Strom et al., 1992).

The scaling of crater sizes between planetary bodies requires knowledge of (1) the flux ratio of impactors at different sizes considering celestial mechanical aspects (**Section 10.10.5**) and (2) how the conditions of crater formation differ from one planet to another (**Section 10.10.3**). These conditions are mainly defined by the mass of the projectiles, impact velocities, impact angles, gravity acceleration, and target properties of the impacted bodies. In the last decades, scaling laws were derived from low-velocity laboratory-scale experiments with the theoretical extrapolation to realistic impact velocities and large projectile diameters (e.g., Holsapple, 1993; Holsapple and Housen, 2007; Holsapple and Schmidt, 1987; Housen and Holsapple, 2003; Schmidt and Housen, 1987). Computing power and numerical models allowed also improving the extrapolation of experimental data to any scale of impact including the basin-forming impacts (Ivanov, 2006, 2008; Ivanov et al., 2010a,b).

Relative ages can be converted into absolute age estimates using a chronology model (exemplified in **Figure 10**), by assuming that (1) the crater-production function (the original projectile SFD) is accurately known and stable with time, (2) the flux of projectiles onto the surfaces is isotropic, and (3) target-property variations across the target body are negligible. Therefore, the crater densities measured on planetary surfaces that have been exposed for the same duration are assumed to be similar for all planetary bodies.

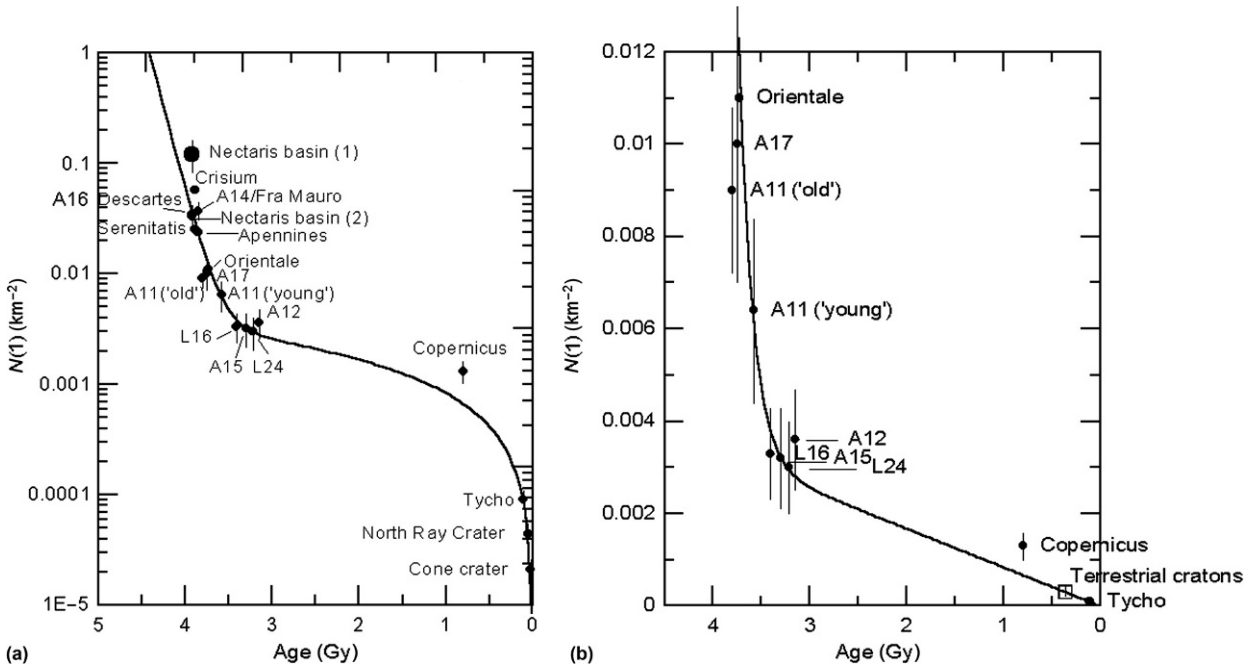


Figure 10 The accumulation rate of lunar impact craters, described as the number of craters with $D = 1$ km, $N(1)$, estimated by fitting real crater counts with a production function, (a) a semilogarithmic plot, (b) in a linear scale for younger crater population. Crater counts and sample ages are from Neukum et al. (1975b), Neukum and Ivanov (1994), Stöffler and Ryder (2001), redrafted after Neukum et al. (2001a) and Table 4 of Stöffler and Ryder (2001).

10.10.7.2 Uncertainties and Issues of the Crater-Based Surface Dating Method

As discussed in Section 10.10.4 and documented in Figures 6, 7, and 9, different authors and groups have developed SFDs for lunar and other planetary cratered surfaces. Two most commonly used varieties of the crater-based age-determination method have been derived by Neukum and Hartmann. The shapes of their crater-production functions differ for crater diameters larger than 1 km and vary significantly for crater diameters between 2 and 25 km (Figure 9; e.g., Ivanov, 2001; Werner and Tanaka, 2011). Fitting differently shaped crater-production functions to the same crater density measurements of the lunar sampling site-related geologic units results in contrasting estimates of the projectile fluxes. As an example, fitting different crater SFDs to the same original measurements of crater superposing the Apollo 16 landing site is shown in Figure 11. The Apollo 16 landing site samples are associated with the Nectaris basin, and thus, its formation age and superposed crater frequencies are considered as one of the key calibration points for the lunar evolution and the cratering chronology models.

Figure 12 shows the effect of different flux estimates at different reference diameters on the derivation of cratering chronology models, while the isotope ages are kept the same for all. The previously described example for the Nectaris basin is marked by a vertical dashed line. Application of these models leads to very diverse interpretations of the (earliest) evolutionary history of the Moon although based on exactly the same observations.

Despite the great variety of models, traditionally, in the crater-based age-determination these models share, one common feature, which is the assumption that the cratering-rate decays smoothly and monotonically. For example, no spike in cratering rate is assumed during the LHB period at about 3.9 Ga (e.g., Baldwin, 1974; Wetherill, 1975). A monotonic extrapolation for times prior about 4.1 Ga (the assumed and debated date for the Nectaris basin formation on the Moon) is used. Restricted by the maximum observed crater density (Figure 12), rarely ages earlier than ca. 4.2 billion years are found, while dated lunar samples suggest solid surface formation as old as 4.47 ± 0.07 Ga (Nyquist et al., 2010). However, also recent Nice models (e.g., Morbidelli et al., 2012b) do not provide such old ages. The shape of the curve may be distorted by considering large and small crater diameters separately, and this is reflected in the diversity of shapes of the various models. Nevertheless, this fact led to the idea that the projectile population changed over time (Fassett et al., 2012a,b; Head et al., 2010; Strom et al., 2005; Werner, 2008), which violates the traditional postulate that the source region of the projectile population, planetary orbits, and the mass distribution within the solar system have never been reconfigured over its lifetime.

Marchi et al. (2012a) argued for the significance of velocity effects for craters around 45–60 km diameter on about 4 Ga old lunar surfaces, resulting from the postulated 'Nice event' during the migration of giant planets. This migration modified the probable source regions of the impacting bodies, and thus with time, the average impact velocities were changed (Bottke et al., 2012; Morbidelli et al., 2012b). Marchi et al. (2012a) considered modern SFD of the MABs and accounted for processes that transported the asteroids from the MAB into the inner solar system. Yet this study is based on Lunar Orbiter Laser Altimeter (LOLA) topography maps, for which the craters below 60 km are not satisfactorily recognized (Figure 13).

Moreover, Marchi et al. (2012a) used the scaling law for transient crater diameters (Schmidt and Housen, 1987) $D_{tc} \sim U^{0.44}$ (eqn [10]), while for complex craters, the crater widening during the collapse stage results in a different velocity exponent $D \sim U^{0.528}$ (eqn [16]). The impact velocity difference proposed by Marchi et al. (2012a) for 'old' (pre-'Nice event')

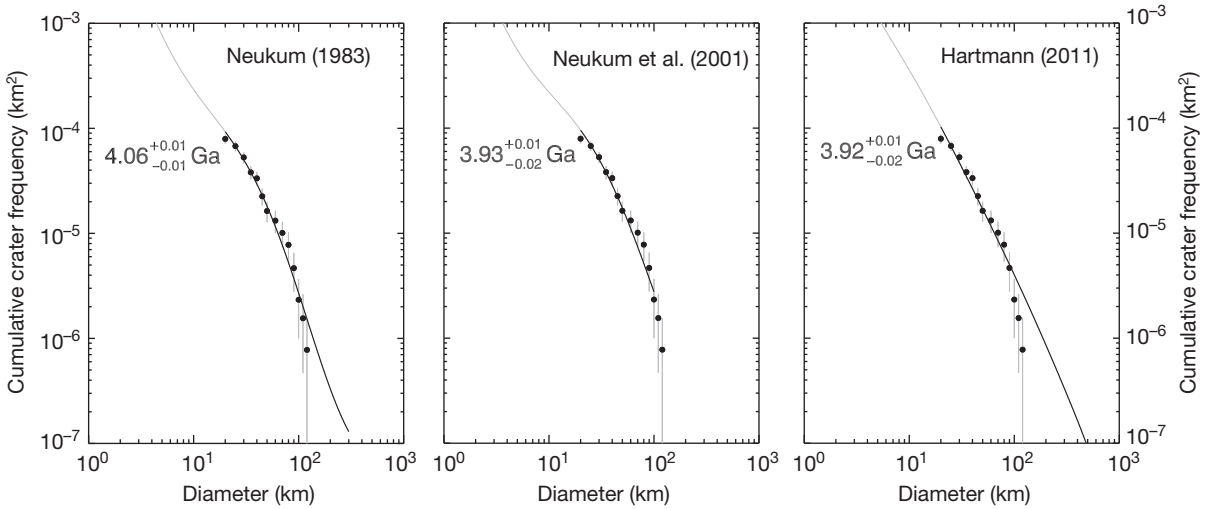


Figure 11 Fit of three different crater-production functions (**Figure 10**) by Neukum (1983), Neukum et al. (2001a,b), and Hartmann (2011, personal communication) to the same measurement of the Nectaris superposed craters (Neukum et al., 1975a,b). The measurement is used for the original calibration of the cratering chronology model by Neukum (1983). The ages (determined all with the same chronology model) reflect the variations in the extrapolated cumulative crater frequency for craters of diameters of 1 km and larger and, therefore, the need for adjusting the chronology model to reproduce the correct isotope age associated with the Nectaris basin formation age. However, additional uncertainties in the isotope age association are so far neglected.

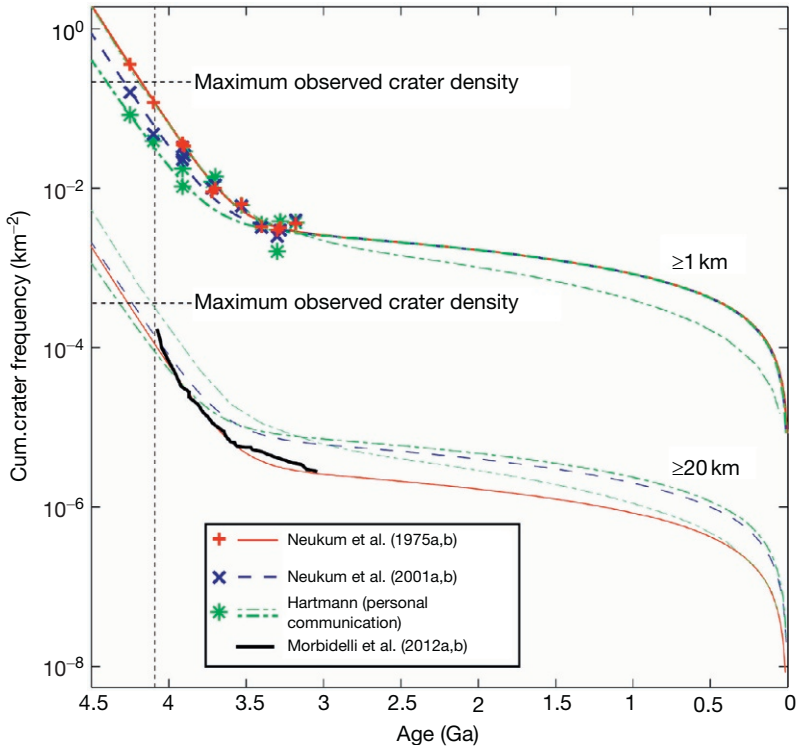


Figure 12 Cratering chronology models derived for four different lunar crater-production function curves (see **Figure 9**) shown for $D \geq 1$ km and ≥ 20 km. The Nice model (Morbidelli et al., 2012c) calibrated and therefore best-fit to the chronology model from Neukum et al. (1975a,b) is compared with the latter (Werner, unpublished). Age variations are depicted in **Figure 11**, if adjustments of the chronology models according to the specific size–frequency are ignored.

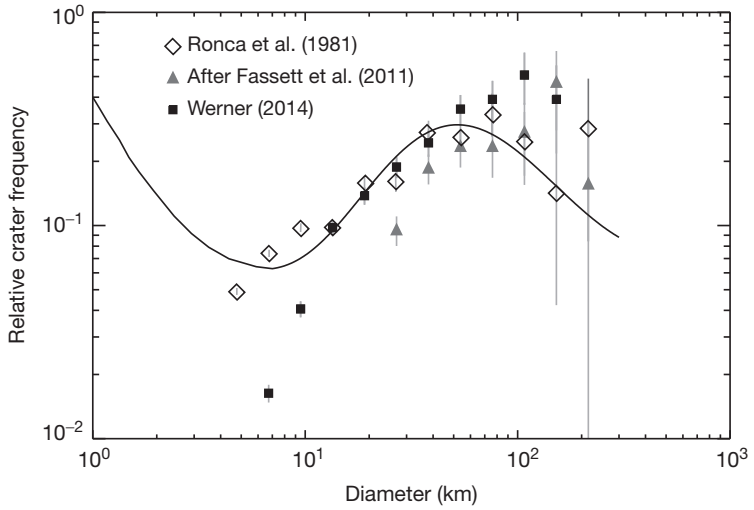


Figure 13 Example of relative (to $N_{cum} \sim D^{-3}$) crater frequencies on the Moon. Counts on LOLA topographic data (triangle) underestimate craters in the range below 60 km compared to images (diamond, square).

and 'young' (post-'Nice event') of a factor of 2 should be of the order of 1.7 to shift the crater diameter for the same projectile from 45 to 60 km. Further investigations are needed to compare observations by Marchi et al. (2012a,b,c) with the celestial modeling results (Bottke et al., 2012), which predicted an increase of median velocity U from 9 to 21 km s^{-1} (the U-ratio $21/9 = 2.33^{0.528}$ should result in D-ratio for complex craters of $2.33^{0.528} \sim 1.56$), and the characteristic lunar crater diameter of 45 km shifts to ~70.4 km, not to 60 km as observed by Marchi et al. (2012a).

As demonstrated in **Figure 11** as well, the flux rate during the period younger than about 3.3 Ga for which constant flux

is suggested also depends on the shape of the crater SFD. Thus, these shape uncertainties leave room for speculations on, for example, the constancy of the bombardment flux. First, there are arguments of catastrophic collision of large MABs, which may have temporarily increased the bombardment flux into planetary-crossing orbits (e.g., Bottke et al., 2005) or dynamically more complex evolution of the solar system, such as outlined by Bottke et al. (2012). Second, there is a hypothesis (based on limited observational data) about the gradual decay of the bombardment flux during the last 3 Ga (Hartmann et al., 2007; Quantin et al., 2007). However, these observations could also be explained by lowering the flux throughout the period of constant flux.

10.10.7.3 Chronology Model Transfer from the Moon to Other Planetary Bodies

Already early on, different approaches on how to transfer the lunar chronology model to other solid surface planetary bodies were published and compared (e.g., Hartmann, 1973, 1977; Hartmann et al., 1981; Neukum and Hiller, 1981; Neukum and Wise, 1976; Soderblom et al., 1974). While for the standard cratering chronology approach, a monotonically and smoothly decaying cratering rate is assumed, competing interpretations suggested a peak in the cratering rate at about 3.9 Ga ago. The correlation of rates on other planetary bodies has been based on maximum crater frequencies (e.g., Neukum and Wise, 1976; Wetherill, 1975) although associated to different absolute age values. More modern approaches rely on the estimated cratering-rate ratios (**Section 10.10.5**). Following the idea of a temporal marker horizon (Wetherill, 1975) that most of the craters including the basins have formed either during the LHB (e.g., Wetherill, 1975) or as the smooth decaying flux of the planetary formation (e.g., Neukum and Wise, 1976, or Hartmann, 1977), it can be tested how the characteristics of the fluxes for the terrestrial bodies compare. Werner (2008) and Fassett and Head (2010) compared basin formation ages for Mars and the Moon and suggested that the basin-forming population disappeared and similarly the high-flux rates for smaller craters. However, the maximum surface age of the Moon seems older than on Mars. Fassett et al. (2011) studied crater frequencies for Mercury and the Moon with a qualitative comparison cratering rates, suggesting that the surface of Mercury is less cratered than the lunar one. Preliminary results by Werner and Ivanov (2012) revealed that general characteristics in the basin-superposing crater frequencies and basin formation ages for Mercury, Mars, and the Moon with the current chronology models are in agreement with each other, but details need further investigation, which require an updated crater scaling for craters and basins larger than about 250 km in diameter. In general, estimates based on asteroid bombardment-rate comparison could constrain these early correlations; examples for Mars and Mercury are Hartmann and Neukum (2001) and Neukum et al. (2001b) but are lacking a comparative test for all three together. Now, Werner (2014) could show for different types of cratering chronology models that appropriate impact rate ratios are crucial for the interpretation of the bombardment history for Mars and Mercury. A detailed and more fundamental description of deriving proper cratering rates for all terrestrial planets is given by Ivanov (2001) and Ivanov et al. (2001). Aspects of this treatment comprise the derivation of (1) the lunar impact velocity distribution, (2) the projectile SFD from the lunar cratering records, (3) the impact probability and impact velocity distribution for the target planet, and (4) the calculation of a crater SFD on the target planet using the lunar-derived projectile SFD:

Step (1). The lunar impact velocity distribution is derived based on probability estimates for near-Earth asteroids with Earth-crossing orbits to impact the Earth–Moon system, resulting in a function f_v to define the probability of impact with a given velocity.

Step (2). This is to determine the projectile SFD from the lunar cratering records in a full treatment, which requires the solution of the integral equation

$$\frac{dN}{dD} = \int_{U_{min}}^{U_{max}} dU \int_{\phi_{min}}^{90^\circ} d\phi \frac{dN}{dD_p} \frac{dD_p}{dD} f_U(U) f_\phi(\phi) \qquad [25]$$

to find the differential SFD for projectiles, dN/dD_p, from the measured SFD for craters, dN/dD. Therefore, the probability of impacts, f_U, for the impact velocity range from U_{min} to U_{max} has to be determined, and from cratering scaling laws, the dependence of the crater diameter D on the projectile size, D_p, impact velocity, U, and impact angle ϕ, so that $D = D(D_p, U, \phi)$, are available. The frequency of oblique impacts, f_ϕ, is known from theoretical considerations (eqn [17]). The current standard of scaling laws is discussed in **Section 10.10.3**.

In a simplified approach, as in the early publications, one can compute the average impact velocity $<U>$ (**Table 3**) and assume that all impacts occur at the most probable impact angle of 45°. In this case, the number of projectiles creating craters of diameter D is the same as the number of projectiles with diameter D_p, provided the scaling law $D = D(D_p, <U>, 45^\circ)$ is known. Hence, $dN/dD_p = dN/dD$. In both the simplified and full-treatment (eqn [25]) approaches, we still assume that all projectiles have the same density. In reality, realistic projectile densities range from about 500 kg m^{-3} (porous comet nucleus) to about 7000 kg m^{-3} (iron meteoroids), including most practically important values of 2700 kg m^{-3} for S-type asteroids and 1300 kg m^{-3} for dark (e.g., C-type) asteroids (Britt et al., 2002; Stuart and Binzel, 2004). However, for the S-type asteroid 25143 Itokawa, a density of only 2100–2300 kg m^{-3} was found (Fujiwara et al., 2006).

Step (3). This is the derivation of impact probability and impact velocity distribution for the target planet, which again requires model assumptions or observational information for the orbits of small bodies crossing the orbit of a target planet. While the average impact velocity estimated by different groups for various terrestrial planets is commonly in agreement (see **Table 3**), the determination of the average impact rate (the number of impacts of bodies of the same size per unit time per unit area of the target planet surface) is limited by the current understanding of orbital dynamics of asteroids and comets. However, the comparison of pioneer publications (Hartmann, 1977; Hartmann et al., 1981; Shoemaker, 1977), which based the estimates on a dozen of small-body orbits known that time, with recent estimates, which include hundreds of orbits, is not dramatically

different in estimates. Starting with observed small-body orbits, one can make estimates of the average impact-rate ratio for a given planet relative to the Moon, as illustrated in Table 3.

Step (4). Finally, a crater SFD for the target planet is calculated using the lunar-derived projectile SFD. For this purpose, cratering scaling laws are applied, which take the surface rock properties of the target planet and its gravity acceleration into account. The full-treatment approach uses eqn [25] to compute the crater SFD. In the simplified treatment, the average impact velocity and an impact angle of 45° are assumed.

In the succeeding text, we illustrate available data on the relative cratering rate for terrestrial planets.

10.10.7.4 The Earth–Moon System

The Earth and the Moon are subjected approximately to the same projectile flux (Table 3), so the Earth–Moon comparison is a good starting point to discuss challenges of interplanetary comparison of cratering rates. For the Moon, we have reasonably reliable crater counts; in a few cases, they are *calibrated* with isotope ages of lunar returned samples. Most available crater counts are for lunar mare surfaces, while crater records for younger surfaces are typically rare.

On Earth, only a very small number of craters are recognized, because of erosive losses of small ones and plate tectonic and subduction activity, which recycles two-thirds of the Earth surface within about 250 Ma (Figure 14). For most craters on Earth, good age estimates from isotope dating of impact melt or by stratigraphic relationships exist. In addition, the data on frequency of meteoroid entries to the terrestrial atmosphere give a snapshot of the modern bombardment rate (Ivanov, 2006).

The Earth–Moon bombardment history starts with the Earth–Moon system formation at about 4.5 Ga ago. Commonly accepted, the Moon has been formed near the Earth's Roche limit at a distance of about three Earth radii after a giant impact (e.g., Cameron and Ward, 1976, and recently reviewed by Canup et al., 2013) and has been moved to the current orbit at about 60 Earth radii by tidal forces (Ruskol, 1966; Touma, 2000). The impact-rate ratio in the Earth–Moon system partially depends on the Earth–Moon distance (Bandermann and Singer, 1973), defining the modest nearside/farside and leading/trailing asymmetry of impact rates decreasing for larger approaching velocity. The severity of this asymmetry depends also on the approach velocities of celestial projectiles. Attempts to reveal a (small) lunar cratering asymmetry are continuing (e.g., Morota and Furumoto, 2003). Recent theoretical efforts have been utilized to improve our understanding of the lunar cratering asymmetry issues (Gallant and Gladman, 2006; Le Feuvre and Wieczorek, 2005, 2006; Morota et al., 2006), but they do not give consistent results. The most robust of these resent studies (Gallant and Gladman, 2006) favor a minor (a few percent) enhancement of the cratering rate in the leading hemisphere of the Moon. Werner and Medvedev (2010) studied the global young rayed lunar crater distribution, which revealed a similar pattern as the predicted spatial asymmetries (Le Feuvre and Wieczorek, 2008, 2011). However, composition and microbombardment and statistical limitations restricted the interpretation (Werner and Medvedev, 2010). At the current level of lunar surface dating uncertainties, we consider that effects such as cratering asymmetries can be neglected.

The comparison of available data for the Earth–Moon system in general results in the conclusion that the modern (10 year-scale) and geologically recent (100–400 Ma) bombardment flux is the same as proposed for the 3 Ga averaged lunar cratering rate within a factor of 2 (Grieve, 1984; Grieve and Pesonen, 1995; Grieve and Shoemaker, 1994; Ivanov, 2006; Oberst et al., 2012; Stuart and Binzel, 2004; Werner et al., 2002). Some details of this comparison (compare Figure 17) point to difference, which could be a systematic misfit with physical meaning. In the succeeding text, we illustrate possible reasons for this and possible ways of the reconciliation.

Figure 15 summarizes observational data of cumulative collision probability of small bodies with projectile diameters larger than D_p for the Earth. Small-projectile diameter data (D_p in the range of 10–20 cm) are based on Earth-based fireball registration (Halliday et al., 1996), and the entry to the terrestrial atmosphere is considered as collision. Data for larger bodies (D_p in the range of 1–10 m) are derived from satellite registration of bolide flashes (Brown et al., 2002), where projectile masses are estimated from the light flash energy and converted to kinetic energy, with the sizes estimated for an average impact velocity of 20 km s^{-1} and average density of 2700 kg m^{-3}. For the large-sized range ($D_p \geq 5$ km), the estimate is derived from the observed NEO population with an albedo distribution for Near Earth asteroids (NEAs) according to a model by Binzel et al. (2004) and Stuart and Binzel (2004). The intermediate-sized crater range of $10\ \text{m} < D_p < 5$ km suffers from the observational incomplete set of observed NEAs, for which the uncertainty increases with decreasing size. A number of population prediction were published (Bottke et al., 2002; D'Abramo et al., 2001; Mainzer et al., 2011a; Rabinowitz et al., 2000; Stuart, 2001; Stuart and Binzel, 2004). However, these predictions also need to assume the albedo, which may change with size and orbital parameters (e.g., Mainzer et al., 2011b; Werner et al., 2002), and orbital parameters, to allow a complete assessment of the averaged population collision probability. (We note that the general update of the 10-year-old debiased NEA model (Bottke et al., 2002) is announced (Granvik et al., 2012, 2013).)

For the evaluation of realistic estimates of impact probability, astronomical and geologic records can be combined. For the compendium of terrestrial impact records shown in Figure 15, scaling laws are used to convert the terrestrial crater diameter into a projectile mass for an assumed average impact velocity and the average density of projectiles. Figure 16 shows the same data for the terrestrial cratering records, plotting the global probability of the formation of impact craters with crater diameter larger than D. The terrestrial cratering probability is directly derived from crater ages and spatial crater density (Grieve, 1984; Grieve and Shoemaker, 1994; Shoemaker and Shoemaker, 1990). Estimates for the NEO population are converted from a kinetic energy frequency model (Stuart and Binzel, 2004) into terrestrial impact crater diameter with scaling laws discussed in the preceding text (using eqns [8], [15a], and [15b]). To some extent, the NEO kinetic energy estimates depend on the assumed albedo

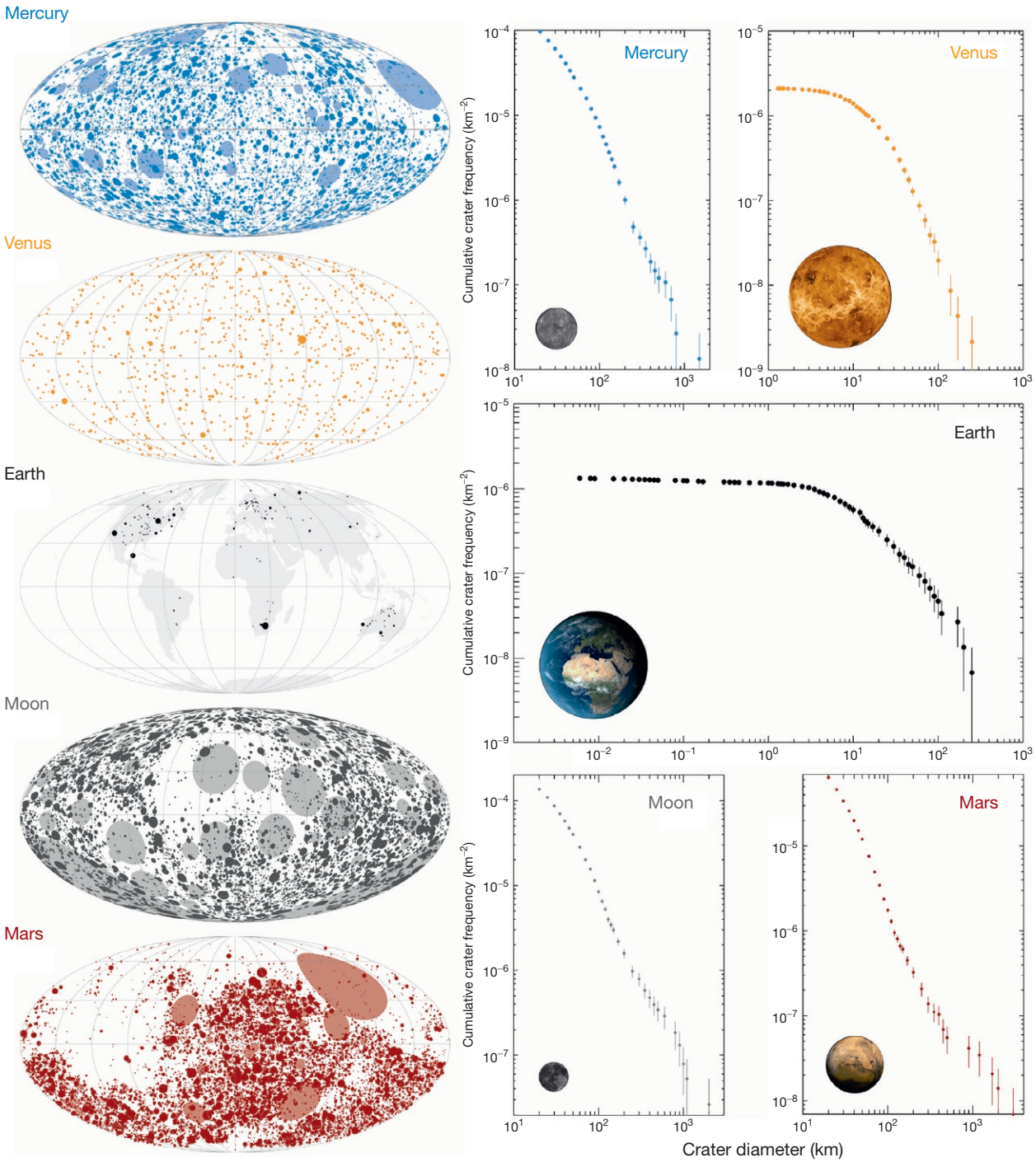


Figure 14 Global spatial crater distribution for the Earth, the Moon, Mercury, Venus, and Mars and the related global crater SFD for the same objects. The distributions for Venus and Earth include all craters, while for Mars, the Moon, and Mercury, only craters larger than 20 km are displayed. Because of geologic activity, as well as atmospheric interaction of the projectiles in the cases of Venus and Earth, all curves significantly deviate from the assumed crater-production function. Catalogs used for deriving the crater distributions are from Head et al. (2010) for the Moon, modified after Barlow (1988) for Mars; Venus' craters are plotted according to Herrick and Phillips (1994); for Earth, we used the Earth Impact Database (http://www.passc.net/EarthImpactDatabase/index.html) and our own database for Mercury.

distribution among NEOs on various orbits. Here, we follow the proportion of *dark* objects of low density and *bright* objects with higher density proposed by Stuart and Binzel (2004). Within a factor of 2, both astronomical (snapshot) and geologic (average for 100–400 Ma) estimates of the terrestrial cratering rate are similar, and the correlation confirms the relative constancy of the bombardment flux to Earth in recent time (<0.5 Ga).

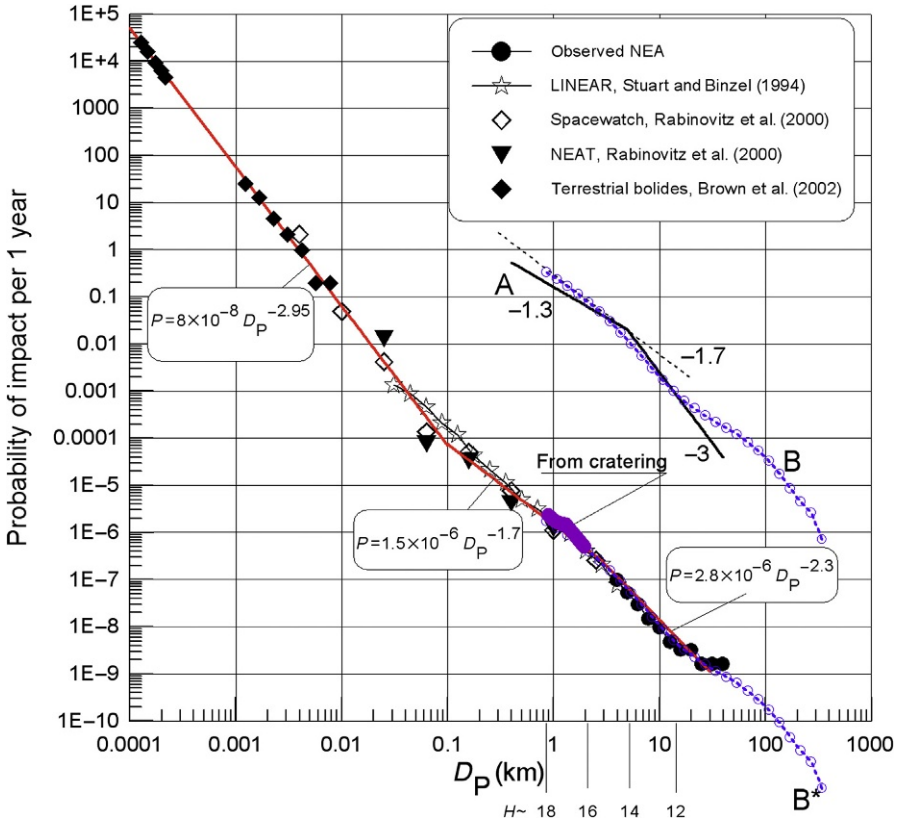


Figure 15 Probability estimates for impacts over the total Earth's surface for small bodies larger than the projectile diameter, D_p. Data for bolides and fireballs ($D_p < 10$ m) are from Halliday et al. (1996) and Brown et al. (2002); observed large near-Earth objects (NEOs) are selected from the 'astorb.dat' file (October 2005, provided through the Minor Planet Center); predictions for small NEOs are derived by Rabinowitz et al. (2000), Stuart (2001), and Stuart and Binzel (2004). Thick purple segment labeled 'From cratering' based on the terrestrial cratering rate over the last 100 Ma converted to D_p for average impact velocity (reproduced from Hughes (2000) A new approach to the calculation of the cratering rate of the Earth over the Last 125 ± 20 Myr. *Monthly Notices of the Royal Astronomical Society* 317: 429–437). Main asteroid belt SFD estimates (plotted here in the arbitrary vertical position) are from Ivezic et al. (2001) in the form of two-segment power law (curve A) with exponent indexes −1.3, −1.7, and −3 proposed by Ivezic et al.(2001) and from Jedicke et al. (2002) (curve B). The latter is scaled to the NEO's probability level (curve B*) to ease the comparison between the SFD of NEOs and main asteroid belts. Absolute magnitude $H < 18$ is given for comparison with the diameter D_p values.

The current bombardment rate derived as global terrestrial probability of impacts of small bodies with an average density of about 2700 kg m^{-3} (bolide and fireball data) and a mix of larger objects (**Figures 15** and **16**) can be analytically described by a simplifying set of power-law segments similarly to the lunar crater-production function of Hartmann (**Section 10.10.4.2**):

$$P(>D_P) \approx 8\times10^{-8} D_P^{-2.95} \quad \text{for } D_P < 100\,\text{m} \qquad [26a]$$

$$P(>D_P) \approx 1.5\times10^{-6} D_P^{-1.7} \quad \text{for } 100\,\text{m} < D_P < 1\,\text{km} \qquad [26b]$$

$$P(>D_P) \approx 2.8\times10^{-6} D_P^{-2.3} \quad \text{for } 1\,\text{km} < D_P < 20\,\text{km} \qquad [26c]$$

Equation [26a]–[26c] respect the general non-power-law behavior of the SFD of asteroids and are constructed to represent all published data within a factor of 2, valid for the different power-law segments, and extrapolation beyond the given ranges is invalid. The terrestrial composite-bombardment flux (eqn [26a]–[26c]) is constructed only to be compared to lunar cratering-rate estimates. For an average impact velocity of 20.9 and 19.2 km s^{-1} for the Earth and the Moon, respectively, and the effective gravity-enhanced cross sections of 7.12×10^8 km^2 and 3.85×10^7 km^2 (Stuart and Binzel, 2004), the Earth–Moon bolide ratio is 1.42. Note the difference with $R_b = 1.68$ estimated from observed orbits of NEA with $H < 17$ (see **Table 3**). Values derived by Stuart and Binzel (2004) for the Earth–Moon impact ratio and average lunar impact velocity are 1.61 and 19.3 km s^{-1}, while Le Feuvre and Wieczorek (2011) found 1.58 and 19.7 km s^{-1}. **Figure 17** shows a comparison of the modern Earth–Moon projectile distribution as derived in the preceding text with lunar production functions (**Sections 10.10.4.2**) in the form of the cumulative number of craters accumulated at the lunar surface per 1 Ga. This comparison is sensitive to the scaling laws (**Section 10.10.3**) used. For example, *nonporous* scaling (eqn [8]) provides a good fit for large craters with diameters $D > 10$ km. The production rate of small craters in regolith ($D < 300$ m) is fitted well with the *porous-rock* (dry sand) scaling (eqn [9]). In the intermediate-sized crater range, *nonporous* scaling overestimates the crater-production rate by an order of magnitude near diameters of about 1 km. If we assume that the upper lunar crust is porous enough in the roughly upper 1 km layer, such a variation in scaling laws and therefore differences in shape of the lunar and terrestrial crater SFD may be realistic

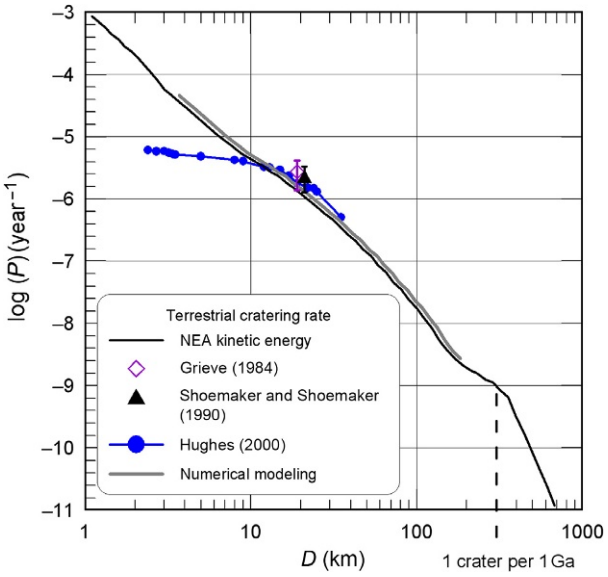


Figure 16 Cratering rate estimates for large terrestrial craters (Grieve, 1984; Shoemaker and Shoemaker, 1990) in comparison with NEOs' impact probability, derived for the population predicted by Stuart and Binzel (2004), using the scaling laws derived from experiments (eqns [8], [15a], and [15b] – black curve) and with the scaling law derived from the numerical modeling of terrestrial craters (eqn [16] – gray curve).

(Ivanov, 2006, 2008), suggesting also that albedo variations with size may not be the only explanation for the discrepancy between lunar crater and estimated asteroid SFDs (Mainzer et al., 2011b; Werner et al., 2002). This was used by Marchi et al. (2009), who interpreted variations in measured crater SFD by means of diameter-range-wise application of different scaling laws, when comparing with modeled asteroid SFDs. More thoroughly, this approach has been tested by numerical modeling of impact crater formation in a hypothetical three-layered lunar crust (Wünnemann et al., 2012), which suggests that the resulting crater diameter can vary up to 20% between impacts into basement rock and regolith, with the most effect near the layer boundaries, where the resulting transient crater cavity diameters become even larger. We note once more that many model results for transient cavity sizes should be verified in the future to make more solid prognoses of the final rim crater diameters.

Reevaluation of published lunar geophysical data (Horvath et al., 1980; Toksoz et al., 1972) by Lognonne et al. (2003) shows longitudinal seismic velocity as low as ~1 km s^{-1} above ~1 km depth below the lunar surface and an increase of 3–4 km s^{-1} to a depth of ~5 km, which is in agreement with and constrains the validity of a *porous* scaling law for lunar impacts applied for a 1 km thick upper layer (Ivanov, 2006). Present results from the GRAIL mission suggest significant density variations for the

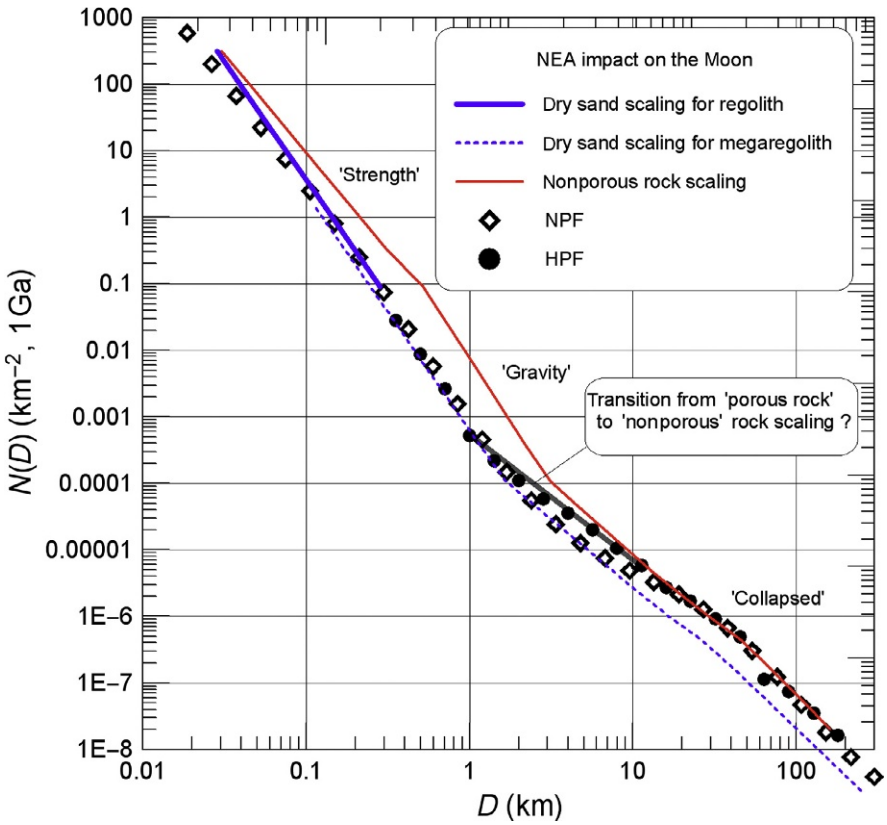


Figure 17 Comparison of the modern Earth–Moon projectile SFD with lunar production functions (**Sections 10.10.4.2**) in the form of the cumulative number of craters accumulated at the lunar surface per 1 Ga and derived with different scaling laws (**Section 10.10.3**).

upper few kilometers of the lunar crust, similarly to seismic observations (Wieczorek et al., 2013).

The spatial distribution and crater size–frequency distribution for the Moon are shown in Figure 14. Highland units are most densely cratered. The interior of most large impact basins was filled after their formation, and the crater density in these mare units is significantly less compared to the highlands.

When calibrating interplanetary surface dating, for example, for Mars and Mercury, porosity variations of the upper subsurface rocks may be different due to different gravity and/or the presence of mechanisms of pore space infilling (e.g., water-caused weathering and massive amounts of ground ice at depths below a few hundred meters). The scaling of crater size D for a given projectile may be different from that on the Moon and can affect the calculated crater-production rate for a specified D.

10.10.7.5 Mercury

Based on the observed asteroid orbit distributions, Neukum et al. (2001b) determined the cratering rate for Mercury relative to the Moon with an average impact velocity of 32 km s^{-1} and R_b of 0.7–0.85. Direct estimates for asteroids with $H<17$ give an average impact velocity of $U\approx35$ km s^{-1} and a bolide ratio of $R_b=0.92$ (Table 3). Scaling laws (eqns [8], [15a], and [15b]) result in the cratering-rate ratios of 1.1 for $D<1$ km and 0.9–1.3 for craters with $1<D$ (km) <100. In the last few years, several additional numerical models and semianalytic interpretations of the cratering-rate ratios have been published for Mercury (e.g., Le Feuvre and Wieczorek, 2008, 2011; Marchi et al., 2009). For these models, they consider layered target properties (Fassett et al., 2009) and size-dependent dynamical evolution of the projectile populations (Marchi et al., 2009, 2013). Although the results differ, depending on which input projectile population is used, detailed scaling laws and methods to determine the collision probabilities are used, and they generally agree within the range of uncertainties. Werner (2014) specifically tested the different models for Mercury in comparison with the Moon, based on basin formation ages, and came to the conclusion that cratering-rate ratios used are not sufficient and especially the model by Marchi et al. (2013) is incoherent with the Nice model assumptions.

The spatial distribution of craters and basins on Mercury is shown in Figure 14, together with the SFD of this observed crater distribution.

10.10.7.6 Venus

According to simple estimates (Table 3), the R_b Venus–Moon bolide ratio is close to 1.1 from observed inner solar system small-body population with $H<17$. This suggests an impact rate about 1.5 times smaller than on Earth, but with higher average impact velocity. First, average age estimates for the Venusian surface were based on radar images from the Russian Venera 15/16 mission (Ivanov and Bazilevskii, 1987; Ivanov et al., 1986) using the approximation of a nearly similar cratering rate on Earth and Venus, followed by Hartmann (1987) who derived lunar-based crater SFD isochrons for Venus. McKinnon et al. (1997) had used small-body orbital set derived from earlier observations (Shoemaker et al., 1990, 1994) to derive the average crater-retention age of Venusian surface as about 700 Ma. Venusian surface-age estimates range between 500 and 1000 Ma.

The global spatial crater distribution is shown in Figure 14, as well as its crater SFD. The crater distribution on Venus globally cannot be distinguished from random (Herrick and Phillips, 1994), so that it is impossible to derive detailed geologic time frames solely based on cratering statistics.

10.10.7.7 Mars

Mars is one of the few solar system bodies for which a group of meteorites has been recognized, the SNC meteorites (Wood and Ashwal, 1981). However, for most the source sites on Mars are unknown (Werner et al., 2013), but ages derived at laboratories on Earth suggest long-lasting geologic, especially volcanic, activity on Mars. Starting with Mariner 9 data, both Hartmann and Neukum (reviewed by Hartmann and Neukum, 2001) have worked on improving the calibration of the crater chronology for Mars (Ivanov, 2001). As for the comparison of the Moon and Mercury, many aspects of crater scaling (Section 10.10.3), the actual shape of the crater SFD (Section 10.10.4), and cratering-rate estimates (Section 10.10.5) impose uncertainties on the comparison. Nevertheless, studies of volcanic activity on Mars (e.g., Hartmann et al., 1981; Neukum and Hiller, 1981; Werner et al., 2009) show a good agreement between the phases of active volcanism on Mars and basaltic Martian meteorite ages. Mars is now also the first planet (after Moon and Earth) for which in-situ radiometric and exposure ages exist (Farley et al. 2014). Thus, the Martian cratering chronology model by Werner et al. (2014) is calibrated in situ by ages measured at Gale crater and Mojave crater, which is suggested to be the source crater for all shergottites (Werner et al. 2014).

Figure 14 shows the Martian global spatial crater distribution, with densely cratered highland units, similar to the lunar ones. The less densely cratered units are the largest impact basins, the almost hemisphere-sized northern lowlands, and all volcanic provinces.

10.10.7.8 Cratering on Asteroids

Several asteroids are imaged, and parts of their cratering record are studied. Asteroids with (partially) measured crater distributions are listed in Table 4. Only for a few asteroids, for

Table 4 A compilation of asteroids for which crater size–frequency distributions are determined

Asteroid	*Sources*
253 Mathilde	Veverka et al. (1999)
243 Ida	Chapman et al. (1996a)
951 Gaspra	Chapman et al. (1996b)
433 Eros	Chapman et al. (2002)
25143 Itokawa	Hirata et al. (2009)
2867 Šteins	Marchi et al. (2010)
21 Lutetia	Marchi et al. (2012b)
4 Vesta	Marchi et al. (2012c)
4179 Toutatis	Zou et al. (2014)

example, for Vesta, it is the class of HED meteorites, meteorites can possibly be linked and thus provide temporal constraints, else calibration of cratering chronologies for asteroids is nearly impossible. The challenge is to find appropriate parameters for the crater scaling to compare the measured crater SFDs with the lunar one. Important parameters are gravity, target properties, and impact velocities to understand crater erasing processes. The irregular shape and unknown heterogeneities of the body add to the scaling uncertainties. The cratering records observed on the surface of asteroids, especially small ones, are cited to discuss the proportion between primary and secondary craters in planetary records, because the primary ejecta probably do not fall back on small asteroids. On the other hand, the surfaces of these small bodies are supposedly as old as the solar system and the surface will probably be saturated with craters at most sizes. The bombardment history of these bodies can influence the asteroid spin rate and axis.

10.10.7.9 Interpretation of the Small-Sized Crater Statistics

Challenges in determining the crater-production function from crater size–frequency measurements are a number of well-known and long-standing difficulties. These difficulties not only are partly inherent issues of the cratering process but also more significantly are related to the interpretation of the crater record and thus made some questions about the overall credibility of the method. Numerous crater-production functions exist, because *the* major challenge in the development of the crater-counting technique and for surface-age determination is to find large, homogeneous, and undisturbed surfaces and to determine the *true* crater-production function. Observed crater distributions could be diverted due to the following:

1. *Average impact velocities*, which compress or stretch the curve along the diameter axis
2. *Target properties*, which change the curve at the small crater range
3. *Admixture of secondary craters*, which steepens the curve
4. *Resurfacing processes*, which shallow the curve
5. *Mapping and counting* (average and distortion), when geologic boundaries are ignored

The shape of the crater-production function must be determined over the largest possible diameter range. However, young units have been used for measurements of the small crater range and old units for large craters in order to derive statistically meaningful measurements from an unmodified cratering record. The shape of the curve may be distorted by considering large and small crater diameters separately, and this is reflected in the diversity of shapes of the various models. Nevertheless, this fact led to the idea that the projectile population may have varied with time (Fassett et al., 2012a,b; Head et al., 2010; Strom et al., 2005; Werner, 2008). These authors have cited various reasons for these apparent or real changes in the SFDs.

In the recent years, with Mars orbiter cameras (HiRISE on the NASA Mars Reconnaissance Orbiter, MOC on the NASA Global Surveyor mission, and HRSC on the ESA Mars Express) and Thermal Emission Imaging System (THEMIS on the NASA Mars Odyssey mission) showing craters as small as $D = 1$ m on various geologic formations of various scales, there has been increasing emphasis on the question of whether the small craters can be used to date the formation and evolution of young and small-scale features, such as formations with glacier-like morphology, debris aprons, landslide deposits, mantled deposits, and exhumed surfaces. Almost 10 years of Mars imaging accumulated sets of image pairs of the same area and allowed Malin et al. (2006) to collect information about 20 new craters that appeared in 21×10^6 km^2 of Martian surface between May 1999 and March 2006. Similar work has been repeated by Daubar et al. (2013) with HiRISE imagery. Both studies find that the crater formation is lower than extrapolated from the currently available crater-chronology models (Hartmann, 2005; Hartmann and Neukum, 2001). All original lunar crater SFD prototypes are not defined for this very small crater range, which can now be observed on the Moon and Mars. Crater scaling laws and the thin Martian atmosphere will modify the projectile–crater relation between the Moon and Mars significantly. Also, Ivanov (2001) pointed out that the cratering rate is varying with respect to the actual eccentricity of the Martian orbit, which is different at the moment compared to the average over the past 3 billions of years (**Table 3**). All these aspects have not been evaluated in detail, but (1) the results from the observed cratering rate at less than 10 m crater diameter and (2) the extrapolation from crater-chronology models found and calibrated for kilometer-sized craters give some additional confidence to the entire system of the terrestrial planet's surface age in comparison with the cratering statistics.

10.10.7.10 Effects of Target Properties

Explosion experiments on the Earth suggest that target properties may influence crater diameters as well as the shape, even for intermediate crater (see detailed discussions in **Section 10.10.3**). The same projectile can form differently sized craters depending on the target strength, and thus, the slope of the crater SFD may vary if the target is distinctly layered. If, for example, on Mars wind-blown material accumulates above lava plains, craters in the upper dusty layer are larger than in the more competent target material (lava). Due to the cumulative effect of cratering, the observed crater SFD function appears steeper. This mostly affects craters below 200 m in diameter (e.g., Dundas et al., 2010). However, the discrepancy between estimated asteroid and lunar crater SFD (see **Figure 17**) could indicate that this effect is relevant for much thicker layers and affect larger diameter ranges.

10.10.7.11 Secondary Cratering

The basic principle of crater dating of planetary surfaces is the assumption that older surfaces have higher crater densities per unit area than young surfaces. In reality, on a geologically active planet, many processes such as erosion and deposition may modify the crater record partially, making the dating (both absolute and relative) ambiguous in a specific situation. Not only uncertainties in cratering rates, scaling laws, and therefore the analytic description of the crater SFD exist, but also the ratio of crater formation rate from celestial small bodies (primary impacts) to cratering from *secondary* fragments ejected from primary craters is uncertain. The limited

observational constraints to determine the small crater range rate (**Section 10.10.6.3**) keep the discussion alive over the past 50 years (Block and Barlow, 2005; Fielder, 1962; Hartmann, 2005; Ivanov, 2006; König, 1977; McEwen et al., 2005; Shoemaker, 1965; Vickery, 1986; Werner et al., 2009).

Secondary craters on Earth found around large-scale chemical and nuclear detonation craters have been discussed by Roberts (1964). For the largest excavation events, those secondary craters reach diameters of ~40 m. A few reports about secondary craters around natural terrestrial impact craters are not convincing enough, leaving room for further studies (Poag, 1995, 1999). In all cases, secondary craters have been found not far from the source and assumed velocity of ejection is in the range of a few 100 m s^{-1} and below. However, tektite strewn fields on Earth and crater ejecta–rays spreading over half of the Moon (Fielder, 1962) and discovered on Mars in thermal IR imagery (Tornabene et al., 2006) suggest that portions of the ejecta can reach large distances, associated with crater clusters (Bierhaus et al., 2001, 2005; McEwen and Bierhaus, 2006).

10.10.7.11.1 SFD for secondary craters

Shoemaker (1965) proposed that secondary craters have a very steep slope SFD with a cumulative relation of frequency and diameter of $N \sim D^{-m}$ and $m \approx 4$ (in contrast to $m \leq 3$ for small primary craters; compare **Section 10.10.4.1**). Shoemaker (1965) gave two examples for the SFD of secondary craters. Their incremental size–frequency curves are concave and the local exponent m slowly decreases with the crater diameter. This is a typical behavior for SFD of rock fragments formed under pressure (grinding and cracking). Decades of engineering experience resulted in many forms of fragment SFD description, and one most commonly used SFD form is the Weibull distribution (Weibull, 1951). Its application for explosive rock fragmentation is reviewed in detail by Grady and Kipp (1987). The Weibull distribution in its cumulative form expresses the volume of fragments larger than a given size in a fragment population analyzed:

$$V(>x) = V_0 \exp\left[-\left(x/x^*\right)^n\right] \qquad [27]$$

where V_0 is the total volume of fragments, x^* expresses the characteristic size in a population of fragments, and n is the Weibull exponent. According to Grady and Kipp (1987), the Weibull exponent varies from about $n = 6$ for a few fragmentation acts to 1 for heavily comminuting processes, such as a long-term fragmentation in rock mills. For explosion fragments, n ranges typically from 1 to 3. The characteristic feature of the Weibull distribution is the *rollover* at smaller sizes – the distribution for fine fragments is much shallower than for the largest fragments. Ivanov (2006) demonstrated that it is a valid approximation for fragments ejected from the primary crater, which create the secondary craters. Hence, smaller members in a cluster of secondary craters probably are less numerous than it is widely suggested from the steep slope for larger

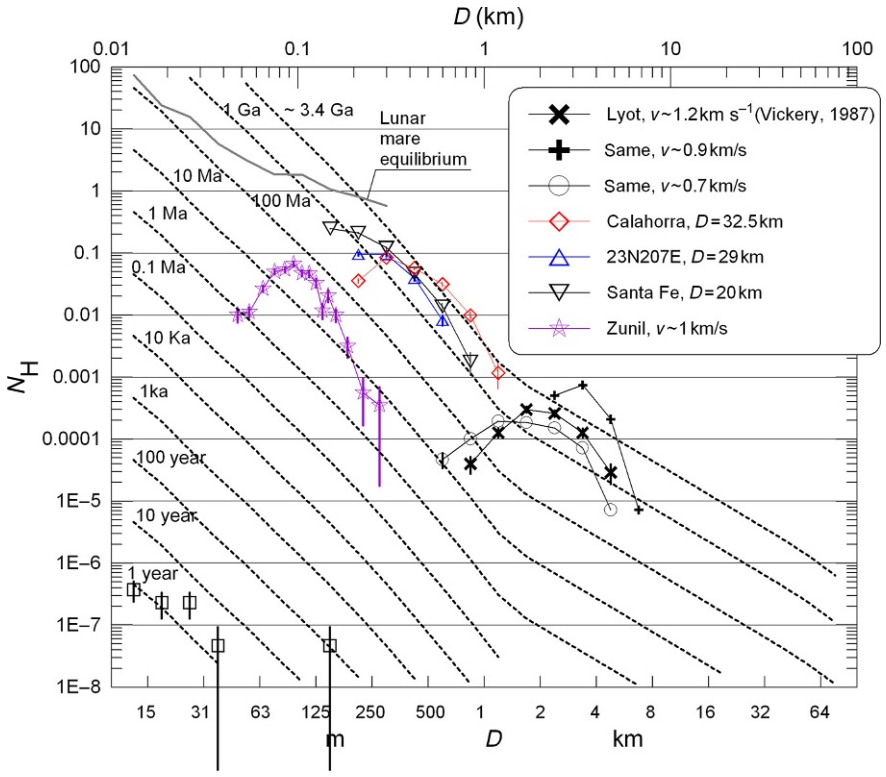


Figure 18 Secondary crater counts (Ivanov, 2006; Werner et al., 2006, 2009) show in incremental plots a deviation from the predicted crater SFD and are of bell-shaped behavior. In comparison, the first direct estimates of small impact crater-production rate on Mars (open squares with error bars; Malin et al., 2006) are shown.

fragments. The rollover character of the crater SFD in clusters is demonstrated by Ivanov (2006) for lunar and Martian secondary craters. The available SFDs for close and remote secondary craters on Mars are shown in **Figure 18** and include one of the highly debated cases, Zunil, a young rayed crater in the Cerberus Plains on Mars. As McEwen's numerical data for SFD of secondaries are unavailable to date, we use HRSC crater counts (Werner et al., 2006) in the area about 300 km west of Zunil, which suggests corresponding ejection or landing fragment velocity of about 1 km s^{-1}. It is interesting to note that fragments that created secondaries around the much larger Lyot crater (also Martian crater) in the same velocity range also have a *bell-shaped* SFD (Ivanov, 2006).

It has been speculated that secondary cratering could impede reproducible age determination (e.g., McEwen et al., 2005). However, one Zunil-sized primary crater probably forms on Mars every 1 Ma, on average. Consistent with this is that the secondary SFD measured in the HRSC image is well bracketed with the 1 and 10 Ma model-age isochrons. The secondary crater densities rise above the 1 Ma isochron at this distance from Zunil, but at larger distances, they drop below (according to McEwen's model). These facts suggest that the Hartmann derivative of isochrons lies closer to Martian realty on surfaces as young as a few Ma than suggest by McEwen et al. (2005). This test has been repeated for all craters on Mars, which appear young due to their prominent ejecta–ray system. All of these passed the test (Hartmann et al., 2010). A more general test, of course, is to try to take into account the total accumulation rate of secondaries from all sizes of primaries (not just Zunil-sized craters), which has been performed by Werner et al. (2009), where they conclude that the contribution of secondaries globally should be less than 10–15%. We note, however, that primary craters of $D < 3$ km are too small to launch large distance secondaries (Head et al., 2002). Issues were raised about the ratio of secondaries to primaries among different-sized craters (e.g., McEwen et al., 2005). It is usually accepted (Shoemaker, 1965) that the ratio is low in the *shallow branch* of the lunar SFD, at $D > 2$ km, and Werner et al. (2009) estimated that the contribution of *random distant background* secondary cratering to the smaller crater *steep branch* distribution segment can be only up to 10–15%. Uncertainties of the actual shape of the crater SFD remain. Yet dated returned samples from Mars do not exist, so arguments in the discussion about error bars in crater-retention age are mostly based on models.

10.10.7.12 Effects of Atmosphere and Geologic Processes

The effects of atmosphere are most obvious in the crater record of Venus (**Figure 14**), where craters smaller than a few kilometers in diameter are completely lacking because of the burnup in the atmosphere. While Mercury and the Moon have no atmosphere, Mars has one. For Mars, however, the significance at present is not well understood. Malin et al. (2006) documented that some of the newly formed impact events are clustered. Ivanov et al. (2011) used these clusters and those newly found in CTX and HiRISE imagery to determine breakup altitudes and effective meteoroid densities, which appear lower than expected. Therefore, they concluded that not only the target properties but also the projectile properties need to be considered, so that the crater–projectile scaling complexity is added.

Especially on Mars, where high-resolution imagery is available, the role of geologic modification needs to be considered, such as aeolian and seasonal processes related to frost and freezing. They modify the small-scale cratering record already within decades.

10.10.8 Conclusions

The exogenic dynamics of planets is a fast-developing system of knowledge, integrating facts and models from many fields of science from celestial mechanics to physics of shock waves. Visible on all planetary bodies, the scars of impact – impact craters – vividly illustrate that exchange among solid bodies is an active agent of the solar system and planetary surface evolution. Individual giant impacts appreciably change the heat balance of planets, producing *impact hot spots* for post-impact magmatic activity. SFDs of numerous small impacts, recorded in crater populations, create an opportunity to estimate not only the age of planetary surfaces but also the rates and efficiency of erosion and deposition processes through studies of degradation and losses of small craters. SFDs of craters can be compared with SFDs of asteroids and comets, resulting in observational constraints for the general solar system evolutionary history. Many topics of exogenic dynamics are still far from fully understood, and many problems are only qualitatively solved. Progress in the exogenic dynamics offers better understanding of planetary formation and evolution.

Acknowledgments

The authors are grateful to W. K. Hartmann for his contributions to the first edition of this chapter. Similarly, we like to thank G. Neukum for his fruitful discussions at many occasions. We acknowledge the efforts of Kai Wünnemann and an anonymous reviewer to improve the manuscript. SCW acknowledges support from the Research Council of Norway through its Centres of Excellence funding scheme, project number 223272 (CEED). BAI is supported by Program 22, Russian Academy of Sciences.

References

Andre SL and Watters TR (2006) Depth to diameter measurements of Mercurian mature complex craters. In: *37th Lunar and Planetary Science Conference*, #2054.

Andre SL, Watters TR, and Robinson MS (2005) The long wavelength topography of Beethoven and Tolstoj basins, Mercury. *Geophysical Research Letters* 32: 21202.

Arnold JR (1965) The origin of meteorites as small bodies. Part II: The model. *Astrophysical Journal* 141: 1536–1547.

Artemieva NA and Ivanov BA (2001) Numerical simulation of oblique impacts: Impact melt and transient cavity size. In: *32nd Lunar and Planetary Science Conference*, #1321.

Arthur DWG, Agnieray AP, Horvath RA, Wood CA, and Chapman CR (1963) The system of lunar craters, quadrant I. *Communications of the Lunar and Planetary Laboratory* 3: 30–60.

Arthur DWG, Agnieray AP, Horvath RA, Wood CA, and Chapman CR (1965) The system of lunar craters, quadrant II. *Communications of the Lunar and Planetary Laboratory* 3: 1–60.

Arthur DWG, Agnieray AP, Pellicori RA, Wood CA, and Weller T (1965) The system of lunar craters, quadrant III. *Communications of the Lunar and Planetary Laboratory* 3: 61–146.

Arthur DWG, Pellicori RA, and Wood CA (1966) The system of lunar craters, quadrant IV. *Communications of the Lunar and Planetary Laboratory* 5: 1–208.

Arvidson R, Boyce J, Chapman C, et al. (1979) Standard techniques for presentation and analysis of crater size-frequency data. *Icarus* 37: 467–474.

Baker DMH and Head JW (2013) New morphometric measurements of craters and basins on Mercury and the Moon from MESSENGER and LRO altimetry and image data: An observational framework for evaluating models of peak-ring basin formation. *Planetary and Space Science* 86: 91–116.

Baldwin RB (1964) Lunar crater counts. *Astronomical Journal* 69: 377–392.

Baldwin RB (1971) On the history of lunar impact cratering: The absolute time scale and the origin of planetesimals. *Icarus* 14(1): 36–52.

Baldwin RB (1974) Was there a "terminal lunar cataclysm" $3.9–4.0 \times 10^9$ years ago? *Icarus* 23: 157–166.

Bandermann LW and Singer SF (1973) Calculation of meteoroid impacts on Moon and Earth. *Icarus* 19: 108–113.

Barlow NG (1988) Crater size-frequency distributions and a revised Martian relative chronology. *Icarus* 75: 285–305.

Barr AC and Canup RM (2010) Origin of the Ganymede–Callisto dichotomy by impacts during the late heavy bombardment. *Nature Geoscience* 3: 164–167.

Bertoldi F, Altenhoff W, Weiss A, Menten KM, and Thum C (2006) The trans-neptunian object UB313 is larger than Pluto. *Nature* 439: 563–564.

Bierhaus EB, Chapman CR, and Merline WJ (2005) Secondary craters on Europa and implications for cratered surfaces. *Nature* 437: 1125–1127.

Bierhaus EB, Chapman CR, Merline WJ, Brooks SM, and Asphaug E (2001) Pwyll secondaries and other small craters on Europa. *Icarus* 153: 264–276.

Binzel RP, Rivkin AS, Stuart JS, Harris AW, Bus SJ, and Burbine TH (2004) Observed spectral properties of near-Earth objects: Results for population distribution, source regions, and space weathering processes. *Icarus* 170: 259–294.

Bloch MR, Fechtig H, Genter W, Neukum G, and Schneider E (1971) Meteorite impact craters, crater simulations, and the meteoroid flux in the early solar system. *Proceedings of the Second Lunar Science Conference* 3: 2639–2652.

Block KM and Barlow NG (2005) Secondary cratering rates on the basaltic plains of Mars and the Moon. In: *36th Lunar and Planetary Science Conference*, #1816.

Borovicka J, Koten P, Spurny P, Bocek J, and Stork R (2005) A survey of meteor spectra and orbits: Evidence for three populations of Na-free meteoroids. *Icarus* 174: 15–30.

Bottke J, William F, Durda DD, et al. (2005) Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* 179(1): 63–94.

Bottke WF and Greenberg R (1993) Asteroidal collision probabilities. *Geophysical Research Letters* 20: 879–881.

Bottke J, William F, Morbidelli A, et al. (2002) Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* 156(2): 399–433.

Bottke WF, Vokrouhlický D, Minton D, et al. (2012) An Archaean heavy bombardment from a destabilized extension of the asteroid belt. *Nature* 485: 78–81.

Bottke WF, Vokrouhlický D, and Nesvorný D (2007) An asteroid breakup 160 Myr ago as the probable source of the K/T impactor. *Nature* 449: 48–53.

Bottke WF, Vokrouhlicky D, Rubincam DP, and Nesvorny D (2006) The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annual Review of Earth and Planetary Sciences* 32: 157–191.

Brasser R (2013) The formation of Mars: Building blocks and accretion time scale. *Space Science Reviews* 174: 11–25.

Brasser R, Walsh KJ, and Nesvorny D (2013) Constraining the primordial orbits of the terrestrial planets. *Monthly Notices of the Royal Astronomical Society* 433: 3417–3427.

Britt D, Yeomans TD, Housen K, and Consolmagno G (2002) Asteroid density, porosity, and structure. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 485–500. Tucson, AZ: University of Arizona Press.

Brown ME (2012) The compositions of Kuiper belt objects. *Annual Review of Earth and Planetary Sciences* 40(1): 467–494.

Brown P, Spalding RE, ReVelle DO, Tagliaferri E, and Worden SP (2002) The flux of small near-Earth objects colliding with the Earth. *Nature* 420: 294–296.

Broz M, Morbidelli A, Bottke WF, Rozehnal J, Vokrouhlický D, and Nesvorný D (2013) Constraining the cometary flux through the asteroid belt during the late heavy bombardment. *Astronomy and Astrophysics* 551: 117.

Buckingham E (1914) On physically similar systems: Illustrations of the use of dimensional equations. *Physical Review* 4: 345–376.

Burchell MJ and Whitehorn L (2003) Oblique incidence hypervelocity impacts on rock. *Monthly Notices of the Royal Astronomical Society* 341: 192–198.

Cameron AGW and Ward WR (1976) The Origin of the Moon. In: *Abstracts for Lunar and Planetary Science Conference 7th*, 120–122.

Canup RM, Barr AC, and Crawford DA (2013) Lunar-forming impacts: High-resolution SPH and AMR-CTH simulations. *Icarus* 222(1): 200–219.

Capek D and Vokrouhlicky D (2004) The YORP effect with finite thermal conductivity. *Icarus* 172: 526–536.

Ceplecha Z (1987) Geometric, dynamic, orbital and photometric data on meteoroids from photographic fireball networks. *Bulletin of the Astronomical Institutes of Czechoslovakia* 38: 222–234.

Chambers JE (2010) Planetesimal formation by turbulent concentration. *Icarus* 208: 505–517.

Chambers JE and Wetherill GW (2001) Planets in the asteroid belt. *Meteoritics and Planetary Science* 36: 381–399.

Chapman CR and Haefner RR (1967) A critique of methods for analysis of the diameter-frequency relation for craters with special application to the Moon. *Journal of Geophysical Research* 72: 549.

Chapman CR and McKinnon WB (1986) Cratering of planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 492–580. Tucson, AZ: University of Arizona Press.

Chapman CR, Merline WJ, Thomas PC, et al. (2002) Impact history of Eros: Craters and boulders. *Icarus* 155: 104–118.

Chapman CR, Ryan EV, Merline WJ, et al. (1996) Cratering on Ida. *Icarus* 120: 77–86.

Chapman CR, Veverka J, Belton MJS, Neukum G, and Morrison D (1996) Cratering on Gaspra. *Icarus* 120: 231–245.

Charnoz SB, Morbidelli A, Dones L, and Salmon J (2009) Did Saturn's rings form during the Late Heavy Bombardment? *Icarus* 199: 413–428.

Chesley SR, Ostro SJ, Vokrouhlicky D, et al. (2003) Direct detection of the Yarkovsky effect by radar ranging to asteroid 6489 Golevka. *Science* 302(5651): 1739–1742.

Cintala MJ and Grieve RAF (1998a) Erratum: Scaling impact melting and crater dimensions: Implications for the lunar cratering record. *Meteoritics and Planetary Science* 33: 1343.

Cintala MJ and Grieve RAF (1998b) Scaling impact-melt and crater dimensions: Implications for the lunar cratering record. *Meteoritics and Planetary Science* 33: 889–912.

Collins GS, Elbeshausen D, Davison TM, Robbins SJ, and Hynek BM (2011) The size-frequency distribution of elliptical impact craters. *Earth and Planetary Science Letters* 310(1): 1–8.

Collins GS, Kenkmann T, Osinski GR, and Wünnemann K (2008) Mid-sized complex crater formation in mixed crystalline-sedimentary targets: Insight from modeling and observation. *Meteorology and Planetary Science* 43: 1955–1977.

Collins GS, Melosh HJ, Morgan JV, and Warner MR (2002) Hydrocode simulations of Chicxulub Crater collapse and peak-ring formation. *Icarus* 157: 24–33.

Collins GS, Wünnemann K, Artemieva N, and Pierazzo E (2013) Numerical modelling of impact processes. In: Osinski GR and Pierazzo E (eds.) *Impact Cratering: Processes and Products*, pp. 254–270. Hoboken, NJ: Wiley.

Croft SK (1981) The excavation stage of basin formation: A qualitative model. In: Schultz PH and Merrill RB (eds.) *Multiring Basins*, pp. 207–225. New York: Pergamon Press.

Croft SK (1985) The scaling of complex craters. *Proceedings of the 15th Lunar and Planetary Science Conference*, *Journal of Geophysical Research* 90: C828–C842.

Crosby A and McKenzie D (2005) Measurements of the elastic thickness under ancient lunar terrain. *Icarus* 173: 100–107.

D'Abramo G, Harris AW, Boattini A, Werner SC, Harris AW, and Valsecchi GB (2001) A simple probabilistic model to estimate the population of near-Earth asteroids. *Icarus* 153: 214–217.

Daubar IJ, McEwen AS, and Byrne (2013) How accurately can we date recent climate change on Mars? In: *Lunar and Planetary Science Conference*, #1719.

Davison TM, Collins GS, Elbeshausen D, Wünnemann K, and Kearsley A (2011) Numerical modeling of oblique hypervelocity impacts on strong ductile targets. *Meteoritics and Planetary Science* 46(10): 1510–1524.

Dence MR (1965) The extraterrestrial origin of Canadian craters. *Annals of the New York Academy of Sciences* 123: 941–969.

Dienes JK and Walsh JM (1970) Theory of impact: Some general principles and the method of Eulerian codes. In: Kinslow R (ed.) *High Velocity Impact Phenomena*, pp. 45–104. New York: Academic Press.

Dones L, Gladman B, Melosh HJ, Tonks WB, Levison HF, and Duncan M (1999) Dynamical lifetimes and final fates of small bodies: Orbit integrations vs Öpik calculations. *Icarus* 142: 509–524.

Dundas CM, Keszthelyi LP, Bray VJ, and McEwen AS (2010) Role of material properties in the cratering record of young platy-ridged lava on Mars. *Geophysical Research Letters* 37: 12203.

Elbeshausen D, Wünnemann K, and Collins GS (2009) Scaling of oblique impacts in frictional targets: Implications for crater size and formation mechanisms. *Icarus* 204: 716–731.

Elbeshausen D, Wünnemann K, and Collins GS (2013) The transition from circular to elliptical impact craters. *Journal of Geophysical Research: Planets* 118(11): 2295–2309. http://dx.doi.org/10.1002/2013JE004477.

Farinella P, Froeschle C, Froeschle C, et al. (1994) Asteroids falling onto the Sun. *Nature* 371: 315–317.

Farnocchia D, Chesley SR, Chodas PW, et al. (2013) Yarkovsky-driven impact risk analysis for asteroid (99942) Apophis. *Icarus* 224: 192–200.

Farley KA, Malespin C, Mahaffy P, et al. (2014) The MSL Science Team In Situ Radiometric and Exposure Age Dating of the Martian Surface. *Science* 343(6169). http://dx.doi.org/10.1126/science.1247166.

Fassett CI, Head JW, Baker DMH, et al. (2012) Large impact basins on Mercury: Global distribution, characteristics, and modification history from MESSENGER orbital data. *Journal of Geophysical Research (Planets)* 117. http://dx.doi.org/10.1029/2012JE004154.

Fassett CI and JW Head (2010) Conditions on Early Mars: Scenarios, Transitions and Events. In: *Lunar and Planetary Science Conference* 41, abs. #1951.

Fassett CI, Head JW, Blewett DT, et al. (2009) Caloris impact basin: Exterior geomorphology, stratigraphy, morphometry, radial sculpture, and smooth plains deposits. *Earth and Planetary Science Letters* 285: 297–308.

Fassett CI, Head JW, Kadish SJ, et al. (2012) Lunar impact basins: Stratigraphy, sequence and ages from superposed impact crater populations measured from Lunar Orbiter Laser Altimeter (LOLA) data. *Journal of Geophysical Research* 117: E00H06. http://dx.doi.org/10.1029/2011JE003951.

Fassett CI, Kadish SJ, Head JW, Solomon SC, and Strom RG (2011) The global population of large craters on Mercury and comparison with the Moon. *Geophysical Research Letters* 38: L10202. http://dx.doi.org/10.1029/2011GL047294.

Fassett CI and Minton DA (2013) Impact bombardment of the terrestrial planets and the early history of the Solar System. *Nature Geoscience* 6(7): 520–524.

Fernandez JA, Gallardo T, and Brunini A (2002) Are there many inactive Jupiter-family comets among the near-Earth asteroid population? *Icarus* 159: 358–368.

Fernandez JA and Ip W-H (1984) Some dynamical aspects of the accretion of Uranus and Neptune: The exchange of orbital angular momentum with planetesimals. *Icarus* 58(1): 109–120.

Fielder G (1962) Ray elements and secondary-impact craters on the Moon. *Astrophysical Journal* 135: 632–637.

French BM and Short NM (eds.) (1968) *Shock Metamorphism of Natural Materials*. Baltimore, MD: Mono Book Corporation, p. 644.

Fujiwara A, Kawaguchi J, Uesugi K, et al. (2006) Global properties of 25143 Itokawa observed by Hayabusa. In: *37th Lunar and Planetary Science Conference*, #1575.

Gallant J and Gladman B (2006) Lunar cratering asymmetries. In: *37th Lunar and Planetary Science Conference*, abstract no. 2336.

DE Gault, WL Quaide and VR Obcrbeck (1968) Impact cratering mechanics and structures. In: BM French and NM Short (eds.) *Shock Metamorphism of Natural Materials*, pp. 87–99. Baltimore, MD: Mono Book Co.

Gault DE (1970) Saturation and equilibrium conditions for impact cratering on the lunar surface: Criteria and implications. *Radio Science* 5: 273–291.

Gault DE and Wedekind JA (1978) Experimental studies of oblique impact. In: *Proceedings of the 9th Lunar and Planetary Science Conference*, pp. 3843–3875. New York: Pergamon Press.

Gerasimov MV, Ivanov BA, Yakovlev OI, and Dikov YP (1999) Physics and chemistry of impacts. In: Ehrenfreund P, Krafft C, Kochan H, and Pirronello V (eds.) *Laboratory Astrophysics and Space Research*, pp. 209–259. Boston, MA: Kluwer Academic Publishers.

Gilbert GK (1893) The Moon's face: A study of the origin of its features. *Philosophical Society of Washington Bulletin* VII: 241–292.

Gladman BJ, Burns JA, Duncan MJ, and Levison HF (1995) The dynamical evolution of lunar impact ejecta. *Icarus* 118: 302–321.

Gladman B, Michel P, and Froeschle C (2000) The near-Earth object population. *Icarus* 146: 176–189.

Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Science* 435: 466–469.

Grady DE and Kipp ME (1987) Dynamic rock fragmentation. In: Atkinson BK (ed.) *Fracture Mechanics of Rock*, pp. 429–475. San Diego, CA: Academic Press.

Granvik M, Morbidelli A, Bottke W, et al. (2012) Source regions for near-Earth objects. In: *AAS/Division for Planetary Sciences Meeting Abstracts*, abstract #305.03.

Granvik M, Morbidelli A, Jedicke R, et al. (2013). A new population model of the orbits and absolute magnitudes of near-earth objects. In: *AAS/Division for Planetary Sciences Meeting Abstracts*, abstract #106.02.

Green DWE (2006) (134340) Pluto, (136199) Eris, and (136199) Eris I (Dysnomia). International Astronomical Union Circular, 8747, p. 1.

Greenberg R and Hartmann WK (1977) Impact strength: A fundamental parameter of collisional evolution (abstract). *Bulletin of the American Astronomical Society* 9: 455.

Greenberg R, Hartmann WK, Chapman CR, and Wacker JF (1978) The accretion of planets from planetesimals. In: Gehrels T and Matthews MS (eds.) *Protostars and Planets: Studies of Star Formation and of the Origin of the Solar System*, pp. 599–620. Tucson, AZ: University of Arizona Press.

Grieve RA (1984) The impact cratering rate in recent time. *Journal of Geophysical Research* 89(supplement): 403.

Grieve RAF and Cintala MJ (1992) An analysis of differential impact melt-crater scaling and implications for the terrestrial impact record. *Meteoritics* 27: 526–538.

Grieve RAF, Dence MR, and Robertson PB (1977) Cratering processes: As interpreted from the occurrence of impact melts. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering*, pp. 791–814. New York: Pergamon Press.

Grieve RAF and Garvin JB (1984) A geometric model for excavation and modification at terrestrial simple craters. *Journal of Geophysical Research* 89: 11561–11572.

Grieve RAF and Pesonen LJ (1995) Terrestrial impact craters: Their spatial and temporal distribution and impacting bodies. *Earth, Moon, and Planets* 71: 357–376.

Grieve RAF and Shoemaker EM (1994) The record of the past impacts on Earth. In: Gehrels T (ed.) *Hazards Due to Comets and Asteroids*, pp. 417–462. Tucson, AZ: University of Arizona Press.

Halliday I, Griffin AA, and Blackwell AT (1996) Detailed data for 259 fireballs from the Canadian camera network and inferences concerning the influx of large meteoroids. *Meteoritics and Planetary Science* 31: 185–217.

Hansen BMS (2009) Formation of the terrestrial planets from a narrow annulus. *Astrophysical Journal* 703: 1131–1140.

Harris AW (2003) The impact frequency of near-Earth asteroids. In: *AAS/Division of Dynamical Astronomy Meeting*, vol. 34, abstract #02.04.

Hartman WK (1978) Martian cratering V – Toward an empirical Martian chronology, and its implications. *Geophysical Research Letters* 5: 450–452.

Hartmann WK (1969) Terrestrial, lunar, and interplanetary rock fragmentation. *Icarus* 10: 201–213.

Hartmann WK (1970a) Note: Lunar cratering chronology. *Icarus* 13: 299.

Hartmann WK (1970b) Note: Preliminary note on lunar cratering rates and absolute time-scales. *Icarus* 12: 131.

Hartmann WK (1972) Inter-planet variations in scale of crater morphology Earth, Mars, Moon. *Icarus* 17: 707.

Hartmann WK (1973) Martian cratering 4: Mariner 9 initial analysis of cratering chronology. *Journal of Geophysical Research* 78: 4096–4116.

Hartmann WK (1977) Relative crater production rates on planets. *Icarus* 31: 260–276.

Hartmann WK (1984) Does crater 'saturation equilibrium' occur in the solar system? *Icarus* 60: 56–74.

Hartmann WK (1999) Martian cratering VI. Crater count isochrons and evidence for recent volcanism from Mars Global Surveyor. *Meteoritics and Planetary Science* 34: 167–177.

Hartmann WK (2005) Martian cratering 8: Isochron refinement and the chronology of Mars. *Icarus* 174: 294–320.

Hartmann WK, Anguita J, de la Casa MA, Berman DC, and Ryan EV (2001) Martian cratering. Part 7: The role of impact gardening. *Icarus* 149: 37–53.

Hartmann WK, Farinella P, Vokrouhlicky D, et al. (1999) Reviewing the Yarkovsky effect: New light on the delivery of stone and iron meteorites from the asteroid belt. *Meteoritics and Planetary Science* 34: 161–167.

Hartmann WK and Kuiper GP (1962) Concentric structures surrounding lunar basins. *Communications of the Lunar and Planetary Laboratory* 1: 51–66.

Hartmann WK and Neukum G (2001) Cratering chronology and evolution of Mars. In: Kallenbach R, Geiss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 165–194. Dordrecht: Kluwer Academic Press.

Hartmann WK, Quantin C, and Mangold N (2007) Possible long-term decline in impact rates. Part 2: Lunar impact-melt data regarding impact history. *Icarus* 186: 11–23.

Hartmann WK, Quantin C, Werner SC, and Popova O (2010) Do young martian ray craters have ages consistent with the crater count system? *Icarus* 208: 621–635.

Hartmann WK, Ryder G, Dones L, and Grinspoon D (2002) The time-dependent intense bombardment of the primordial Earth/Moon system. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 493–512. Tucson, AZ: University of Arizona Press.

Hartmann WK, Strom RG, Weidenschilling SJ, et al. (1981) Chronology of planetary volcanism by comparative studies of planetary craters. In: *Basaltic Volcanism on the Terrestrial Planets*, pp. 1050–1127. Elmsford, NY: Pergamon Press.

Hartmann WK and Wood CA (1971) Moon: Origin and evolution of multiring basins. *Moon* 3: 3–78.

Head JW III, Fassett CI, Kadish SJ, et al. (2010) Global distribution of large lunar craters: Implications for resurfacing and impactor populations. *Science* 329(5998): 1504–1507.

Head JN, Melosh HJ, and Ivanov BA (2002) Martian meteorite launch: High-speed ejecta from small craters. *Science* 298: 1752–1756.
Herrick RR and Phillips RJ (1994) Implications of a global survey of Venusian impact craters. *Icarus* 111: 387–416.
Hiesinger H and Head JW III (2002) Topography and morphology of the Argyre basin, Mars: Implications for its geologic and hydrologic history. *Planetary and Space Science* 50: 939–981.
Hiesinger H, Head JW, Wolf U, Jaumann R, and Neukum G (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108: 5065. http://dx.doi.org/10.1029/2002JE001985.
Hiesinger H, Jaumann R, Neukum G, and Head JW (2000) Ages of mare basalts on the lunar nearside. *Journal of Geophysical Research* 105: 29239–29276.
Hirata N, Barnouin-Jha OS, Honda C, et al. (2009) A survey of possible impact structures on 25143 Itokawa. *Icarus* 200: 486–502.
Holsapple KA (1993) The scaling of impact processes in planetary sciences. *Annual Review of Earth and Planetary Sciences* 21: 333–373.
Holsapple KA and Schmidt RM (1979) A material-strength model for apparent crater volume. In: Lunar Science Institute (ed.) *Proceedings of the 10th Lunar and Planetary Science Conference*, pp. 2757–2777. New York: Pergamon.
Holsapple KA and Schmidt RM (1980) On the scaling of crater dimensions. Part 1: Explosive processes. *Journal of Geophysical Research* 85: 7247–7256.
Holsapple KA and Schmidt RM (1982) On the scaling of crater dimensions. Part II: Impact processes. *Journal of Geophysical Research* 87: 1849–1870.
Holsapple KA and Schmidt RM (1987) Point source solutions and coupling parameters in cratering mechanics. *Journal of Geophysical Research* 92: 6350–6376.
Holsapple KA and Housen KR (2007) A crater and its ejecta: An interpretation of Deep Impact. *Icarus* 187: 345–356.
Horvath P, Latham GV, Nakamura Y, and Dorman HJ (1980) Lunar near-surface shear wave velocities at the Apollo landing sites as inferred from spectral amplitude ratios. *Journal of Geophysical Research* 85: 6572–6578.
Housen KR and Holsapple KA (2003) Impact cratering on porous asteroids. *Icarus* 163: 102–119.
Hughes DW (2000) A new approach to the calculation of the cratering rate of the Earth over the Last 125 ± 20 Myr. *Monthly Notices of the Royal Astronomical Society* 317: 429–437.
Ivanov BA (1981) Cratering mechanics. *Advances in Science and Technology of VINITI. Ser. Mechanics of Deformable Solids,* pp. 60–128. Moscow: VINITI Press, English translation: NASA Tech. Memorandum 88477/N87–15662, 1986.
Ivanov BA (2001) Mars/Moon cratering rate ratio estimates. In: Kallenbach R, Geiss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 87–104. Dordrecht: Kluwer Academic Press.
Ivanov BA (2004) Heating of the lithosphere during meteorite cratering. *Solar System Research* 38: 266–279.
Ivanov BA (2005) Numerical modeling of the largest terrestrial meteorite craters. *Solar System Research* 39: 381–409.
Ivanov BA (2006) Notes about secondary crater SFD. In: Workshop on Surface Ages and Histories: Issues in Planetary Chronology, 23 May 2006, pp. 28–29. Huston, TX: Lunar and Planetary Institute, LPI Contribution No. 1320.
Ivanov BA (2008) Geologic effects of large terrestrial impact crater formation. In: Adushkin VV and Nemchinov IV (eds.) *Catastrophic Events Caused by Cosmic Objects*, p. 163. Berlin: Springer, ISBN: 978-1-4020-6451-7.
Ivanov BA and Artemieva NA (2001) Transient cavity scaling for oblique impacts. In: *32nd Lunar and Planetary Science Conference*, #1327.
Ivanov BA and Artemieva NA (2002) Numerical modeling of the formation of large impact craters. In: Koeberl C and MacLeod KG (eds.) *Catastrophic Events and Mass Extinctions: Impact and Beyond*, pp. 619–630. Boulder, CO: Geological Society of America, Special Papers 356.
Ivanov BA and Bazilevskii AT (1987) Comparison of crater retention ages on Earth and Venus. *Astronomicheskii Vestnik* 21: 136–143.
Ivanov BA, Bazilevskii AT, Krivchkov VP, and Chernaia IM (1986) Impact craters of Venus – Analysis of Venera 15 and 16 data. *Journal of Geophysical Research* 91: 413.
Ivanov BA and Kamyshenkov D (2012a) Impact cratering: Scaling law and thermal softening. In: *Lunar and Planetary Institute Science Conference Abstracts*, vol. 43, abs. 1407. http://www.lpi.usra.edu/meetings/lpsc2012/pdf/1407.pdf.
Ivanov BA, Melosh HJ, and McEwen A (2010b) New small impact craters in high resolution HiRISE images – III. In: *Lunar and Planetary Science Conference*, #1533.
Ivanov B, Melosh HJ, and McEwen A (2011) New impact craters and meteoroid densities on Mars. In: *EPSC-DPS Joint Meeting*, p. 786.
Ivanov BA, Melosh HJ, and Pierazzo E (2010c) Basin-forming impacts: Reconnaissance modeling. In: Gibson RL and Reimold WU (eds.) *Large Meteorite Impacts and Planetary Evolution IV*, pp. 29–49. Boulder, CO: Geological Society of America, Special Papers 465.
Ivanov BA, Nemchinov IV, Svetsov VA, Provalov AA, Khazins VM, and Phillips RJ (1992) Impact cratering on Venus: Physical and mechanical models. *Journal of Geophysical Research* 97: 16167–16181.
Ivanov BA, Neukum G, Bottke WF Jr., and Hartmann WK (2002) The comparison of size-frequency distributions of impact craters and asteroids and the planetary cratering rate. *Asteroids* III: 89–101.
Ivanov BA, Neukum G, and Wagner R (2001) Size-frequency distributions of planetary impact craters and asteroids. In: Marov MY and Rickman H (eds.) *Collisional Processes in the Solar System*, pp. 1–34. Dordrecht: Kluwer Academic Publishers.
Ivanov BA and Stöffler D (2005) The Steinheim impact crater, Germany: Modeling of a complex crater with central uplift. In: *36th Lunar and Planetary Science Conference*, abstract #1443.
Ivezic Z, Tabachnik S, Rafikov R, et al. (2001) Solar system objects observed in the Sloan Digital Sky Survey commissioning data. *Astronomical Journal* 122: 2749–2784.
Jedicke R, Larsen J, and Spahr T (2002) Observational selection effects in asteroid surveys. In: Bottke WF, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 71–87. Tucson, AZ: University of Arizona Press.
Joeleht A, Kirsimae K, Plado J, Versh E, and Ivanov B (2005) Cooling of the Kardla impact crater. Part II: Impact and geothermal modeling. *Meteoritics and Planetary Science* 40: 21–33.
Kenkmann T, Collins GS, and Wünnemann K (2013) The modification stage of crater formation. In: *Impact Cratering: Processes and Products*, pp. 60–75. Hoboken, NJ: Wiley-Blackwell.
Kinslow R (ed.) (1970) *High Velocity Impact Phenomena*. New York: Academic Press, p. 579.
Kirchoff MR and Schenk P (2010) Impact cratering records of the mid-sized, icy saturnian satellites. *Icarus* 206: 485–497.
Kleine T, Touboul M, Bourdon B, et al. (2009) Hf–W chronology of the accretion and early evolution of asteroids and terrestrial planets. *Geochimica et Cosmochimica Acta* 73: 5150–5188.
König A (1977) *Investigations of Primary and Secondary Impact Structures on the Moon and Laboratory Experiments to Study the Ejecta of Secondary Particles*. PhD Thesis, Ruprecht Karl University, Heidelberg, 88 p. (in German).
Le Feuvre M and Wieczorek MA (2005) The asymmetric cratering history of the moon. In: *36th Lunar and Planetary Science Conference*, p. 2043.
Le Feuvre M and Wieczorek MA (2006) The asymmetric cratering history of the terrestrial planets: Latitudinal effect. In: *37th Lunar and Planetary Science Conference*, p. 1841.
Le Feuvre M and Wieczorek MA (2008) Nonuniform cratering of the terrestrial planets. *Icarus* 197(1): 291–306.
Le Feuvre M and Wieczorek MA (2011) Nonuniform cratering of the Moon and a revised crater chronology of the inner Solar System. *Icarus* 214(1): 1–20.
Levison HF, Bottk WF, Gounelle M, Morbidelli A, Nesvorny D, and Tsiganis K (2009) Contamination of the asteroid belt by primordial trans-Neptunian objects. *Nature* 460(7253): 364–366.
Levison HF, Dones L, Chapman CR, Stern SA, Duncan MJ, and Zahnle K (2001) Could the lunar 'Late Heavy Bombardment' have been triggered by the formation of Uranus and Neptune? *Icarus* 151: 286–306.
Levison HF, Morbidelli A, Dones L, Jedicke R, Wiegert PA, and Bottke WF (2002) The mass disruption of Oort cloud comets. *Science* 296(5576): 2212–2215.
Levison HF, Morbidelli A, Tsiganis K, Nesvorny D, and Gomes R (2011) Late orbital instabilities in the outer planets induced by interaction with a self-gravitating planetesimal disk. *Astronomical Journal* 142: 152–162.
Levison HF and Duncan MJ (1997) From the Kuiper belt to Jupiter-family comets: The spatial distribution of ecliptic comets. *Icarus* 127: 13–32.
Lindblad BA, Neslusan L, Porubcan V, and Svoren J (2003) IAU meteor database of photographic orbits – Version 2003. *Earth, Moon, and Planets* 93: 249–260.
Lissauer JJ and Stevenson DJ (2007) Formation of giant planets. In: Reipurth B, Jewitt D, and Keil K (eds.) *Protostars and Planets V*. Tucson, AZ: University of Arizona Press, pp. 591–616, 951 pp.
Lognonne P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: Implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.
Love S and Ahrens TJ (1996) Catastrophic impacts on gravity dominated asteroids. *Icarus* 124: 141–155.
Mainzer A, Grav T, Bauer J, et al. (2011) NEOWISE observations of near-Earth objects: Preliminary results. *Astrophysical Journal* 743(2), article id 156, 17 pp.
Mainzer A, Grav T, Masiero J, et al. (2011) NEOWISE studies of spectrophotometrically classified asteroids: Preliminary results. *Astrophysical Journal* 741(2), article id 90, 25 pp.
Malhotra R (1993) The origin of Pluto's peculiar orbit. *Nature* 365: 819–821.

Malin MC, Edgett KS, Posiolova LV, McColley SM, and Dobrea EZN (2006) Present-day impact cratering rate and contemporary gully activity on Mars. *Science* 314: 1573–1577.

Marchi S, Bottke WF, Kring DA, and Morbidelli A (2012a) The onset of the lunar cataclysm as recorded in its ancient crater populations. *Earth and Planetary Science Letters* 325: 27–38.

Marchi S, Massironi M, Vincent J-B, et al. (2012b) The cratering history of asteroid (21) Lutetia. *Planetary and Space Science* 66(1): 87–95.

Marchi S, McSween HY, O'Brien DP, et al. (2012c) The violent collisional history of asteroid 4 Vesta. *Science* 336(6082): 690–694.

Marchi S, Mottola S, Cremonese G, Massironi M, and Martellato E (2009) A new chronology for the Moon and Mercury. *Astronomical Journal* 137(6): 4936–4948.

Marchi S, Massironi M, Vincent J-B, et al. (2010) The cratering history of asteroid (2867) Steins. *Planetary and Space Science* 58: 1116.

Marchi S, Chapman CR, Fassett CI, Head JW, Bottke WF, and Strom RG (2013) Global resurfacing of Mercury 4.0–4.1 billion years ago by heavy bombardment and volcanism. *Nature* 499: 59–61.

Masiero JR, Mainzer AK, Grav T, Bauer JM, and Jedicke R (2012) Revising the age for the Baptistina asteroid family using WISE/NEOWISE data. *Astrophysical Journal* 14: 759.

Masset F and Snellgrove M (2001) Reversing type II migration: Resonance trapping of a lighter giant protoplanet. *Monthly Notices of the Royal Astronomical Society* 320(4): L55–L59.

McCord TB, Adams JB, and Johnson TV (1970) Asteroid Vesta: Spectral reflectivity and compositional implications. *Science* 168(3938): 1445–1447.

McCrosky RE, Chao C-Y, and Posen A (1978) Prairie network fireball data. Part I: Summary of orbits. Center for Astrophysics Reprint No. 665 (in Russian). *Meteoritka* 37: 44–59.

McEwen AS and Bierhaus EB (2006) The importance of secondary cratering to age constraints on planetary surfaces. *Annual Review of Earth and Planetary Sciences* 34: 540–567.

McEwen AS, Preblich BS, Turtle EP, et al. (2005) The rayed crater Zunil and interpretations of small impact craters on Mars. *Icarus* 176: 351–381.

McKinnon WB and Melosh HJ (1980) Evolution of planetary lithospheres: Evidence from multi-ringed basins on Ganymede and Callisto. *Icarus* 44: 454–471.

McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 969–1014. Tucson, AZ: University of Arizona Press.

Melosh HJ (1989) *Impact Cratering – A Geologic Process*. New York: Oxford, p. 245.

Melosh HJ (2013) The contact and compression stage of impact cratering. In: Osinski GR and Pierazzo E (eds.) *Impact Cratering: Processes and Products*, pp. 32–42. Hoboken, NJ: Wiley-Blackwell.

Melosh HJ and Ivanov BA (1999) Impact crater collapse. *Annual Review of Earth and Planetary Sciences* 27: 385–415.

Melosh HJ and McKinnon W (1978) The mechanics of ringed basin formation. *Geophysical Research Letters* 5: 985–988.

Michael GG, Platz T, Kneissl T, and Schmedemann N (2012) Planetary surface dating from crater size–frequency distribution measurements: Spatial randomness and clustering. *Icarus* 218: 169–177.

Michel P, Froeschle C, and Farinella P (1996) Dynamical evolution of NEAs: Close encounters, secular perturbations and resonances. *Earth, Moon, and Planets* 72: 151–164.

Michel P and Morbidelli A (2013) Population of impactors and the impact cratering rate in the inner solar system. In: Osinski GR and Pierazzo E (eds.) *Impact Cratering: Processes and Products*, pp. 21–31. Hoboken, NJ: Wiley-Blackwell.

Michel P, Morbidelli A, and Bottke WF (2005) Origin and dynamics of near Earth objects. *Comptes Rendus Physique* 6: 291–301.

Michel P and Thomas F (1996) The Kozai resonance for near-Earth asteroids with semi-major axes smaller than 2 AU. *Astronomy and Astrophysics* 307: 310.

Michel P and Yoshikawa M (2005) Earth impact probability of the Asteroid (25143) Itokawa to be sampled by the spacecraft Hayabusa. *Icarus* 179: 291–296.

Milani A, Carpino M, Hahn G, and Nobili AM (1989) Dynamics of planet-crossing asteroids: Classes of orbital behavior: Project SPACEGUARD. *Icarus* 78: 212–269.

Milani A and Farinella P (1995) An asteroid on the brink. *Icarus* 115: 209–212.

Minton DA and Malhotra R (2009) A record of planet migration in the main asteroid belt. *Nature* 457: 1109–1111.

Minton DA and Malhotra R (2010) Dynamical erosion of the asteroid belt and implications for large impacts in the inner Solar System. *Icarus* 207: 744–757.

Mizutani H, Kawakami S-I, Takagi Y, Kumazawa M, and Kato M (1983) Cratering experiments in sands and a trial for general scaling law. *Proceedings of the 13th Lunar and Planetary Science Conference, Part 2*, *Journal of Geophysical Research* 88(supplement): A835–A845.

Moore HJ (1972) Ranger and other impact craters photographed by Apollo 16, pp. 29_45–29_51. *Apollo 16 Preliminary Science Report*. NASA SP-315. Washington, DC: NASA.

Moore JM, Asphaug E, Belton MJS, et al. (2001) Impact features on Europa: Results of the Galileo Europa Mission (GEM). *Icarus* 151: 93–111.

Morbidelli A (2002) *Modern Celestial Mechanics: Aspects of Solar System Dynamics*. New York: Taylor and Francis, p. 356.

Morbidelli A, Brasser R, Gomes R, Levison HF, and Tsiganis K (2010) Evidence from the asteroid belt for a violent past evolution of Jupiter's orbit. *Astronomical Journal* 140: 1391–1401.

Morbidelli A and Gladman B (1998) Orbital and temporal distributions of meteorites originating in the asteroid belt. *Meteoritics and Planetary Science* 33: 999–1016.

Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435(7041): 462–465.

Morbidelli A, Lunine JI, O'Brien DP, Raymond SN, and Walsh KJ (2012) Building terrestrial planets. *Annual Review of Earth and Planetary Sciences* 40(1): 251–275.

Morbidelli A, Marchi S, and Bottke WF (2012) The sawtimeline of the first billon year of lunar bombardment. In: *Workshop on the Early Solar System Bombardment II*, pp. 53–54.

Morbidelli A, Marchi S, Bottke WF, and Kring DA (2012) A sawtooth-like timeline for the first billion years of lunar bombardment. *Earth and Planetary Science Letters* 355: 144–151.

Morbidelli A, Tsiganis K, Crida AL, Levison HF, and Gomes R (2007) Dynamics of the giant planets of the solar system in the gaseous protoplanetary disk and their relationship to the current orbital architecture. *Astronomical Journal* 134: 1790–1798.

Morbidelli A and Vokrouhlicky D (2003) The Yarkovsky-driven origin of near-Earth asteroids. *Icarus* 163: 120–134.

Morgan JV, Warner MR, Collins GS, Melosh HJ, and Christeson GL (2000) Peak-ring formation in large impact craters: Geophysical constraints from Chicxulub. *Earth and Planetary Science Letters* 183: 347–354.

Morota T and Furumoto M (2003) Asymmetrical distribution of rayed craters on the Moon. *Earth and Planetary Science Letters* 206: 315–323.

Morota T, Harujama J, and Furumoto M (2006) Lunar apex-antapex cratering asymmetry and origin of impactors in the Earth–Moon system. In: *37th Lunar and Planetary Science Conference*, p. 1554.

Morrison D, Harris AW, Sommer G, Chapman CR, and Carusi A (2002) Dealing with the impact hazard. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 739–754. Tucson, AZ: University of Arizona Press.

Nesvorný D and Morbidelli A (2012) Statistical study of the early Solar System's instability with four, five, and six giant planets. *Astronomical Journal* 144: 117.

Neukum G (1983) *Meteoritenbombardement und Datierung planetarer Oberflächen*, p. 186. Habilitation Dissertation, Ludwig-Maximilians-University, Munich.

Neukum G and Hiller K (1981) Martian ages. *Journal of Geophysical Research* 86: 3097–3121.

Neukum G and Ivanov BA (1994) Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet, and asteroid cratering data. In: Gehrels T (ed.) *Hazards Due to Comets and Asteroids*, pp. 359–416. Tucson, AZ: University of Arizona Press.

Neukum G, Oberst J, Hoffmann H, Wagner R, and Ivanov BA (2001) Geologic evolution and cratering history of Mercury. *Planetary Space Science* 49: 1507–1521.

Neukum G, Ivanov BA, and Hartmann WK (2001a) Cratering records in the inner solar system in relation to the lunar reference system. In: Kallenbach R, Geiss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 55–86. Dordrecht: Kluwer Academic Press.

Neukum G, Koenig B, and Arkani-Hamed J (1975) A study of lunar impact crater size-distributions. *Moon* 12: 201–229.

Neukum G, Koenig B, Fechtig H, and Storzer D (1975b) Cratering in the earth-moon system – Consequences for age determination by crater counting. *Lunar Science Conference Proceedings*, 3, pp. 2597–2620. New York: Pergamon Press (A78-46741 21-91).

Neukum G, Schneider E, Mehl A, et al. (1972) Lunar craters and exposure ages derived from crater statistics and solar flare tracks. In: *Proceedings of the Lunar Science Conference*, vol. 3, pp. 2793–2810.

Neukum G and Wise DU (1976) Mars – A standard crater curve and possible new time scale. *Science* 194: 1381–1387.

Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research: Planets* 109: 08002.

Nowka D, Wünnemann K, Collins GS, and Elbeshausen D (2010) Scaling of impact crater formation on planetary surfaces. *Abstracts for European Planetary Science Congress 2010*, # 87. http://meetingorganizer.copernicus.org/EPSC2010/EPSC2010-87.pdf.

Nyquist LE, Shih C-Y, Reese YD, et al. (2010) Lunar crustal history recorded in lunar anorthosites. In: *Lunar and Planetary Science Conference*, #1533.
O'Brien DP and Greenberg R (2005) The collisional and dynamical evolution of the main-belt and NEA size distributions. *Icarus* 178: 179–212.
O'Keefe JD and Ahrens TJ (1993) Planetary cratering mechanics. *Journal of Geophysical Research* 98: 17011–17028.
O'Keefe JD and Ahrens TJ (1999) Complex craters: Relationship of stratigraphy and rings to impact conditions. *Journal of Geophysical Research* 104: 27091–27104.
O'Keefe JD, Stewart ST, Lainhart ME, and Ahrens TJ (2001) Damage and rock-volatile mixture effects on impact crater formation. *International Journal of Impact Engineering* 26: 543–553.
Oberst J, Christou A, Suggs R, et al. (2012) The present-day flux of large meteoroids on the lunar surface – A synthesis of models and observational techniques. *Planetary and Space Science* 74(1): 179–193.
Oberst J, Molau S, Heinlein D, et al. (1998) The 'European Fireball Network': Current status and future prospects. *Meteoritics and Planetary Science* 33: 49–56.
Onorato PIK, Uhlmann DR, and Simonds CH (1978) The thermal history of the Manicouagan impact melt sheet, Quebec. *Journal of Geophysical Research* 83: 2789–2798.
Öpik EJ (1951) Stellar models with variable composition. II. Sequences of models with energy generation proportional to the 15th power of temperature. *Proceedings of the Royal Irish Academy, Section A* 54: 44–47.
Öpik EJ (1960) The frequency of crater diameters in Mare Imbrium. *Astronomical Journal* 65: 55.
Öpik EJ (1976) *Interplanetary Encounters: Close-Range Gravitational Interactions*. Amsterdam, Elsevier p. 155.
Osinski GR and Pierazzo E (eds.) (2013) *Impact Cratering: Processes and Products*. Hoboken, NJ: Wiley-Blackwell, 520 pp.
Pierazzo E and Melosh HJ (1999) Hydrocode modeling of Chicxulub as an oblique impact event. *Earth and Planetary Science Letters* 165: 163–176.
Pierazzo E and Melosh HJ (2000a) Understanding oblique impacts from experiments, observations, and modeling. *Annual Review of Earth and Planetary Sciences* 28: 141–167.
Pierazzo E and Melosh HJ (2000b) Hydrocode modeling of oblique impacts: The fate of the projectile. *Meteoritics and Planetary Science* 35(1): 117–130.
Pierazzo E, Vickery AM, and Melosh HJ (1997) A reevaluation of impact melt production. *Icarus* 127: 408–423.
Pieters CM, Tompkins S, Head JW, and Hess PC (1997) Mineralogy of the mafic anomaly in the South Pole-Aitken Basin: Implications for excavation of the lunar mantle. *Geophysical Research Letters* 24: 1903–1906.
Pike RJ (1977a) Apparent depth/apparent diameter relation for lunar craters. In: Lunar Science Institute (ed.) *Proceedings of the 8th Lunar and Planetary Science Conference*, pp. 3427–3436. New York: Pergamon Press.
Pike RJ (1977b) Size-dependence in the shape of fresh impact craters on the moon. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering: Planetary and Terrestrial Implications*, pp. 489–509. New York: Pergamon Press.
Pike RJ (1980) Control of crater morphology by gravity and target type – Mars, Earth, Moon. In: Biberian J-P (ed.) *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 2159–2189. New York: Pergamon Press, International Conference on Cold Fusion.
Poag CW (1995) Late Eocene star wars: The Toms Canyon and Chesapeake Bay impact craters, US East Coast. *Meteoritics* 30: 562.
Poag CW (1999) Secondary craters from the Chesapeake Bay impact. *Lunar and Planetary Science Conference* 30: 1047.
Poelchau MH, Kenkmann T, Thoma K, Hoerth T, Dufresne A, and Schäfer F (2013) The MEMIN research unit: Scaling impact cratering experiments in porous sandstones. *Meteoritics and Planetary Science* 48(1): 8–22.
Popova O, Nemchinov I, and Hartmann WK (2003) Bolides in the present and past martian atmosphere and effects on cratering processes. *Meteoritics and Planetary Science* 38(6): 905–925.
Potts LV and von Frese RRB (2003a) Comprehensive mass modeling of the Moon from spectrally correlated free-air and terrain gravity data. *Journal of Geophysical Research: Planets* 108d, 5-1–5-27.
Potts LV and von Frese RRB (2003b) Crustal attributes of lunar basins from terrain-correlated free-air gravity anomalies. *Journal of Geophysical Research: Planets* 108e, 3-1–3-21.
Potts LV, von Frese RR, Leftwich TE, Taylor PT, Shum CK, and Li R (2004) Gravity-inferred crustal attributes of visible and buried impact basins on Mars. *Journal of Geophysical Research: Planets* 109: 09009.
Quantin C, Mangold N, Hartmann WK, and Allemand P (2007) Possible long-term decline in impact rates. Part 1: Martian geological data. *Icarus* 186: 1–10.
Rabinowitz DL, Helin E, Lawrence K, and Pravdo S (2000) A reduced estimate of the number of kilometer-sized near-Earth asteroids. *Nature* 403: 165–166.
Richardson JE (2009) Cratering saturation and equilibrium: A new model looks at an old problem. *Icarus* 204: 697–715.
Roberts WA (1964) Secondary craters. *Icarus* 3: 348–364.
Roddy DJ, Pepin RO, and Merrill RB (eds.) (1977) *Impact and Explosion Cratering*. New York: Pergamon Press, p. 1301.
Ronca LB, Basilevsky AT, Krychov VP, and Ivanov BA (1981) Lunar craters evolution and meteoroidal flux in pre-mare and post-mare times. *Moon and Planets* 24: 209–229.
Rubincam DP (1998) Yarkovsky thermal drag on small asteroids and Mars–Earth delivery. *Journal of Geophysical Research* 103: 1725.
Rubincam DP (2000) Radiative spin-up and spin-down of small asteroids. *Icarus* 148: 2–11.
Ruskol EL (1966) On the past history of the Earth–Moon system. *Icarus* 5: 221–227.
Ryan EV and Melosh HJ (1998) Impact fragmentation: From the laboratory to asteroids. *Icarus* 133: 1–24.
Safronov VS (1969) *Evolutsia Doplanetnogo Oblaka I Obrazovanie Zemli i Planet (Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets)*. Moscow: Nauka Press, p. 244 (English translation (1972) NASA Technical Translation F-677, US Department of Commerce, Springfield, VA 22151 and Israel Program for Scientific Translation Jerusalem, 206 pp.).
Schenk PM (2002) Thickness constraints on the icy shells of the galilean satellites from a comparison of crater shapes. *Nature* 417: 419–421.
Schenk PM, Chapman CR, Zahnle K, and Moore JM (2004) Ages and interiors: The cratering record of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 427–456. Cambridge: Cambridge University Press.
Schenk PM and Ridolfi FJ (2002) Morphology and scaling of ejecta deposits on icy satellites. *Geophysical Research Letters* 29, 31-1–31-4.
Schmidt RM (1980) Meteor crater: Energy of formation – Implications of centrifuge scaling. In: Biberian J-P (ed.) *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 2099–2128. New York: Pergamon Press, International Conference on Cold Fusion.
Schmidt RM (1987) Preliminary scaling results for crater rim-crest diameter. *Lunar and Planetary Science Conference* 18: 878–879.
Schmidt RM and Housen KR (1987) Some recent advances in the scaling of impact and explosion cratering. *International Journal of Impact Engineering* 5: 543–560.
Senft LE and Stewart ST (2008) Impact crater formation in icy layered terrains on Mars. *Meteoritics and Planetary Science* 43(12): 1993–2013.
Settle M (1980) The role of fallback ejecta in the modification of impact craters. *Icarus* 42: 1–19.
Sheppard SS (2006) Solar system: A planet more, a planet less? *Nature* 439: 541–542.
Shoemaker EM (1962) Interpretation of lunar craters. In: Kopal Z (ed.) *Physics and Astronomy of the Moon*, pp. 283–359. New York: Academic Press.
Shoemaker EM (1963) Impact mechanics at Meteor Crater, Arizona. In: Middlehurst BM and Kuiper GP (eds.) *The Moon, Meteorites and Comets*, pp. 301–336. Chicago, IL: University of Chicago Press.
Shoemaker EM (1965) Preliminary analysis of the fine structure of the lunar surface in Mare Cognitum, pp. 75–134. Technical Report 32-700, Jet Propulsion Laboratory, CalTech.
Shoemaker EM (1977) Astronomically observable crater-forming projectiles. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering: Planetary and Terrestrial Implications*, pp. 617–628. New York: Pergamon Press.
Shoemaker EM, Hackman RJ, and Eggleton RE (1962) Interplanetary correlation of geologic time. *Advances in the Astronautical Sciences* 8: 70–89.
Shoemaker EM, Hait MH, Swann GA, et al. (1970) Lunar regolith at Tranquillity base. *Science* 167: 452.
Shoemaker EM and Shoemaker CS (1990) Proterozoic impact record of Australia. In: *Abstracts for the International Workshop on Meteorite Impact on the Early Earth*, pp. 47–48. Houston, TX: Lunar and Planetary Institute, LPI Contributions 746.
Shoemaker EM, Weissman PR, and Shoemaker CS (1994) The flux of periodic comets near Earth. In: Gehrels T (ed.) *Hazards Due to Comets and Asteroids*, pp. 313–335. Tucson, AZ: University of Arizona Press.
Shoemaker EM and Wolfe RF (1982) Cratering time scales for the Galilean satellites. In: Morrison D (ed.) *Satellites of Jupiter*, pp. 277–339. Tucson, AZ: University of Arizona Press.
Shoemaker EM, Wolfe RF, and Shoemaker CS (1990) Asteroid and comet flux in the neighborhood of Earth. In: Sharpton VL and Ward PD (eds.) *Global Catastrophes in Earth History*, pp. 155–170. Boulder, CO: Geological Society of America, GSA SP-247.
Soderblom LA and Boyce JM (1972) Relative ages of some near-side and far-side terra plains based on Apollo 16 metric photography, , pp. 29-3. *Apollo 16: Preliminary Science Report*. NASA SP-315, C-3 p. Washington, DC: NASA.

Soderblom LA, Condit CD, West RA, Herman BM, and Kreidler TJ (1974) Martian planetwide crater distributions – Implications for geologic history and surface processes. *Icarus* 22: 239–263.

Spudis PD (1993) *The Geology of Multi-Ring Impact Basins*. Cambridge: Cambridge University Press, p. 263.

Squyres SW, Howell C, Liu MC, and Lissauer JJ (1997) Investigation of crater "saturation" using spatial statistics. *Icarus* 125: 67–82.

Stöffler D, Artemieva NA, Ivanov BA, et al. (2004) Origin and emplacement of the impact formations at Chicxulub, Mexico, as revealed by the ICDP deep drilling at Yaxcopoil-1 and by numerical modeling. *Meteoritics and Planetary Science* 39: 1035–1067.

Stöffler D and Grieve RAF (2007) Impactites. In: Fettes D and Desmons J (eds.) *Metamorphic Rocks: A Classification and Glossary of Terms, Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Metamorphic Rocks*. Cambridge: Cambridge University Press, 256 p.

Stöffler D and Ryder G (2001) Stratigraphy and isotope ages of lunar geologic units: Chronological standard for the inner solar system. *Space Science Reviews* 96: 9–54.

Strom RG, Croft SK, and Barlow NG (1992) The Martian impact cratering record. In: Kieffer HH, Jakovsky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 383–423. Tucson, AZ: University of Arizona Press.

Strom RG, Malhotra R, Ito T, Yoshida F, and Kring DA (2005) The origin of planetary impactors in the inner solar system. *Science* 309(5742): 1847–1850.

Strom RG and Neukum G (1988) The cratering record on Mercury and the origin of impacting objects. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 336–373. Tucson, AZ: University of Arizona Press.

Stuart JS (2001) A near-Earth asteroid population estimate from the LINEAR survey. *Science* 294: 1691–1693.

Stuart JS and Binzel RP (2004) Bias-corrected population, size distribution, and impact hazard for the near-Earth objects. *Icarus* 170: 295–311.

Sugano T and Heki K (2004) Isostasy of the Moon from high-resolution gravity and topography data: Implication for its thermal history. *Geophysical Research Letters* 31: 24703.

Tera F, Papanastassiou DA, and Wasserburg GJ (1974) Isotopic evidence for a terminal lunar cataclysm. *Earth and Planetary Science Letters* 22: 1–21.

Toksoz MN, Press F, Anderson K, et al. (1972) Velocity structure and properties of the lunar crust. *Moon* 4: 490.

Toplis MJ, Mizzon H, Monnereau M, et al. (2013) Chondritic models of 4 Vesta: Implications for geochemical and geophysical properties. *Meteoritics and Planetary Science* 48: 2300–2315.

Tornabene LL, Moersch JE, McSween HY, et al. (2006) Identification of large (2–10 km) rayed craters on Mars in THEMIS thermal infrared images: Implications for possible Martian meteorite source regions. *Journal of Geophysical Research* 111(E10), CiteID E10006.

Touma J (2000) The phase space adventure of the Earth and Moon. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 165–178. Tucson, AZ: University of Arizona Press.

Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Science* 435: 459–461.

Ugalde HA, Artemieva N, and Milkereit B (2005) Magnetization on impact structures – Constraints from numerical modeling and petrophysics. In: Kenkmann T, Horz F, and Deutsch A (eds.) *Large Meteorite Impacts III*, pp. 25–42. Boulder, CO: Geological Society of America, Special Paper 384.

Veverka J, Thomas P, Harch A, et al. (1999) NEAR encounter with asteroid 253 mathilde: Overview. *Icarus* 140: 3–16.

Vickery A (1986) Effect of an impact-generated gas cloud on the acceleration of solid ejecta. *Journal of Geophysical Research* 91: 14139–14160.

Vickery AM (1987) Variation in ejecta size with ejection velocity. *Geophysical Research Letters* 14: 726–729.

Vokrouhlicky D, Broz M, Morbidelli A, Bottke WF, and Nesvorny D (2006) Yarkovsky/YORP chronology of asteroid families. *Icarus* 182: 118–142.

Vokrouhlicky D, Broz M, Morbidelli A, et al. (2006) Yarkovsky footprints in the Eos family. *Icarus* 182: 92–117.

Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475: 206–209.

Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2012) Populating the asteroid belt from two parent source regions due to the migration of giant planets – "The Grand Tack" *Meteoritics and Planetary Science* 47: 1941–1947.

Ward WR (1992) Long-term orbital and spin dynamics of Mars. In: Kieffer HH, Jakovsky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 298–320. Tucson, AZ: University of Arizona Press.

Weibull W (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics* 18: 293–297.

Weissman PR, Bottke WF, and Levison HF (2002) Evolution of comets into asteroids. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 669–686. Tucson, AZ: University of Arizona Press.

Werner SC (2008) The early martian evolution – Constraints from basin formation ages. *Icarus* 195(1): 45–60.

Werner SC (2014) Moon, Mars, Mercury: Basin formation ages and implications for the maximum surface age and the migration of gaseous planets. *Earth and Planetary Science Letters* 400(0): 54–65.

Werner SC, Harris AW, Neukum G, and Ivanov BA (2002) NOTE: The near-Earth asteroid size–frequency distribution: A snapshot of the lunar impactor size–frequency distribution. *Icarus* 156: 287–290.

Werner SC and Ivanov BA (2012) Basin formation and the timing of the migration of the gaseous planets. *American Geophysical Union Meeting*, # P11A-1793.

Werner SC, Ivanov BA, and Neukum G (2006) Secondary cratering and age determination on Mars. In: *37th Lunar and Planetary Science Conference*, #1595.

Werner SC, Ivanov BA, and Neukum G (2009) Theoretical analysis of secondary cratering on Mars and an image-based study on the Cerberus Plains. *Icarus* 200(2): 406–417.

Werner SC and Medvedev S (2010) The Lunar rayed-crater population – Characteristics of the spatial distribution and ray retention. *Earth and Planetary Science Letters* 295(1–2): 147–158.

Werner SC and Tanaka TL (2011) Redefinition of the crater–density and absolute-age boundaries for the chronostratigraphic system of Mars. *Icarus* 215(2): 603–607.

Werner SC, Ody A, and Poulet F (2014) The Source Crater of Martian Shergottite Meteorites. *Science* 343(6177): 1343–1346. http://dx.doi.org/10.1126/science.1247282.

Werner SC, van Gasselt S, and Neukum G (2003) Continual geological activity in Athabasca Valles, Mars. *Journal of Geophysical Research: Planets* 108: 8081.

Wetherill GW (1967) Collisions in the asteroid belt. *Journal of Geophysical Research* 72: 2429–2444.

Wetherill GW (1975) Radiometric chronology of the early solar system. *Annual Review of Nuclear and Particle Sciences* 25: 283–328.

Wetherill GW (1978) Accumulation of the terrestrial planets. In: Gehrels T (ed.) *Protostars and Planets: Studies of Star Formation and of the Origin of the Solar System*, pp. 565–598. Tucson, AZ: University of Arizona Press.

Wetherill GW (1980) Formation of the terrestrial planets. *Annual Review of Astronomy and Astrophysics* 18: 77–113.

Wieczorek MA, Neumann GA, Nimmo F, et al. (2013) The crust of the Moon as seen by GRAIL. *Science* 339(6120): 671–675.

Wieczorek MA and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.

Witaker EA (1972) Artificial lunar impact craters: Four new identifications. *Apollo 16 Preliminary Science Report*. NASA SP-315, pp. 29_39–29_45.

Wood CA and Ashwal LD (1981) Meteorites from Mars: Prospects, problems and implications. *Lunar and Planetary Science Conference* 1197–1199.

Woronow A (1978) A general cratering-history model and its implications for the lunar highlands. *Icarus* 34: 76–88.

Wright SP, Tornabene LL, and Ramsey MS (2013) Remote sensing of impact craters. In: *Impact Cratering: Processes and Products*, pp. 194–210. Hoboken, NJ: Wiley-Blackwell.

Wünnemann K, Collins GS, and Melosh HJ (2006) A strain-based porosity model for use in hydrocode simulations of impacts and implications for transient crater growth in porous targets. *Icarus* 180: 514–527.

Wünnemann K, Collins GS, and Osinski GR (2008) Numerical modelling of impact melt production in porous rocks. *Earth and Planetary Science Letters* 269(3–4): 530–539.

Wünnemann K, Marchi S, Nowka D, and Michel P (2012) The effect of target properties on impact crater scaling and the lunar crater chronology. In: *Lunar and Planetary Science Conference*, #1805.

Wünnemann K, Nowka D, Collins GS, Elbeshausen D, and Bierhaus M (2011) Scaling of impact crater formation on planetary surfaces – Insights from numerical modeling. In: *Proceedings of the 11th Hypervelocity Impact Symposium*, pp. 1–14.

Yakovlev OI and Basilevsky AT (1994) Experimental studies of geochemical aspects of impact cratering. In: Dressler B, Grieve RAF, and Sharpton VL (eds.) *Large Meteorite Impacts and Planetary Evolution*, pp. 73–80. Boulder, CO: Geological Society of America, Geological Society of America Special Paper 293.

Zahnle K, Dones L, and Levison HF (1998) Cratering rates on the Galilean satellites. *Icarus* 136: 202–222.

Zahnle K, Schenk P, Levison H, and Dones L (2003) Cratering rates in the outer Solar System. *Icarus* 163: 263–289.

Zeldovich IB (1956) Gas motion due to a short-time pressure impulse (impact). *Akusticheskii Zhurnal* 21: 28–38.
Zharkov VN and Kozenko AV (1990) On the role of Jupiter in the formation of the giant planets. *Soviet Astronomy Letters* 16: 73.
Zieg MJ and Marsh BD (2005) The Sudbury igneous complex: Viscous emulsion differentiation of a superheated impact melt sheet. *GSA Bulletin* 117: 1427–1450.
Zou X, Lia C, Liua J, Wanga W, Lia H, and Pinga J (2014) The preliminary analysis of the 4179 Toutatis snapshots of the Chang'E-2 flyby. *Icarus* 229: 348–354.
Zuber MT, Solomon SC, Phillips RJ, Smith DE, Tyler GL, Aharonson O, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

This page intentionally left blank

10.11 Water on the Terrestrial Planets

J Helbert and E Hauber, Institute for Planetary Research, DLR, Berlin, Germany
D Reiss, Westfälische Wilhelms-Universität Münster, Münster, Germany

© 2015 Elsevier B.V. All rights reserved.

10.11.1 Introduction 368
10.11.1.1 Geophysical Constraints 368
10.11.1.1.1 Stability of water and ice 368
10.11.1.1.2 Adsorption 370
10.11.2 Observational Evidence 370
10.11.2.1 Spectroscopy, Morphology, and Radar 370
10.11.2.1.1 Visual and infrared spectroscopic features of ice and bound water 370
10.11.2.1.2 Gamma-ray spectroscopy and neutron spectroscopy 371
10.11.2.1.3 Morphology 372
10.11.2.1.4 Radar 372
10.11.3 Water on the Surface 373
10.11.3.1 Valley Networks 373
10.11.3.1.1 Morphology and formation processes 373
10.11.3.1.2 Ages 374
10.11.3.1.3 Source of water and climate 374
10.11.3.2 Fluvial and Lacustrine Sedimentary Processes 375
10.11.3.2.1 The orbiter view 375
10.11.3.2.2 The view of the rovers 376
10.11.3.3 Outflow Channels 376
10.11.3.3.1 Morphology 376
10.11.3.3.2 Morphometry/discharge 376
10.11.3.3.3 Formation 378
10.11.3.4 Oceans 378
10.11.3.4.1 Models involving oceans 379
10.11.3.4.2 Assessment of geomorphic evidence 379
10.11.3.4.3 Tests of ocean hypotheses 380
10.11.3.5 Glacial, Permafrost, and Ground-Ice Features 381
10.11.3.5.1 Glacial landforms 381
10.11.3.5.2 Permafrost and ground-ice landforms 382
10.11.3.6 Polar Caps 384
10.11.3.7 Large Debris Flows 385
10.11.3.8 Gullies 386
10.11.3.8.1 Morphology and Occurrence 386
10.11.3.8.2 Formation 387
10.11.3.8.3 Age 387
10.11.3.8.4 Dune gullies 388
10.11.3.8.5 Present-day activity 388
10.11.3.9 Recurring Slope Lineae 389
10.11.4 Water in Mantle and Crust 389
10.11.4.1 Regolith 390
10.11.4.1.1 Moon and Mercury 390
10.11.4.1.2 Mars 391
10.11.4.2 Mineralogy 392
10.11.4.2.1 Methods 393
10.11.4.2.2 Basalt and dust 393
10.11.4.2.3 Hematite 393
10.11.4.2.4 Phyllosilicates 394
10.11.4.2.5 Sulfates 394
10.11.4.2.6 Carbonates 394
10.11.4.2.7 Opaline silica 395
10.11.4.2.8 Chlorides 395
10.11.4.2.9 Zeolites 395
10.11.4.3 Geophysical Consequences 395

10.11.5 Evolution of Water and Climate 396
10.11.6 Summary and Outlook 397
References 398

10.11.1 Introduction

Water is the second most abundant molecule in our solar system, following only molecular hydrogen. It plays an important role in many planetary processes and has significant importance for the habitability of a planetary body. Despite the importance of water, its complex physical behavior is still only partly understood.

This chapter provides an overview of the current knowledge of water and water-related processes on the terrestrial planets from a geophysical and a morphological point of view. Naturally, Mars plays a featured role here, as it is the only terrestrial planet where we have clear evidence for present and past widespread existence and activity of water on and close to the surface. In addition, strong evidence indicates the presence of water-ice deposits on the Moon and Mercury inside permanently shadowed craters close to the poles of the planet. Spectral evidence indicates widespread presence of water molecules across the lunar surface. As the focus of this chapter is on terrestrial planets, we will not discuss the icy moons of the giant planets, which will be dealt with in Chapter 10.18, or Pluto, which is discussed in Chapter 10.19.

This chapter is organized into six sections. Section 10.11.1 presents a general discussion of the geophysical constraints for the stability of ice and water focusing on the conditions found on the terrestrial planets. This is followed by a discussion of observational methods to detect the presence of water and ice. We then discuss water and ice on the terrestrial planets, starting with water on the surface of the terrestrial planets including a discussion of its morphological effects. From there, we move on to discuss water in the mantle and crust of terrestrial planets and discuss briefly the effects of water in the interior of terrestrial planets in Section 10.11.4. This topic is also covered in Chapter 10.08. Finally, a brief discussion on the evolution of water and climate on the terrestrial planets is included. For a more detailed discussion of the interactions between atmospheres and surfaces of terrestrial planets, we refer to Chapter 10.12. We conclude the chapter with a summary and outlook in Section 10.11.6. This includes a very brief discussion of necessary future measurements. For a more detailed discussion of planetary exploration strategies, we refer to Chapter 10.23.

All terrestrial planets are the topic of this chapter, but in much of the discussion, Mars will take center stage as it is the planet showing the most striking evidence of water activity in the past and even in the present.

10.11.1.1 Geophysical Constraints

10.11.1.1.1 Stability of water and ice

Water exists as a liquid at pressures above 611 hPa and temperatures above 273 K, the triple point of water. Under these conditions, the solid, liquid, and vapor phases coexist. Branching from the triple point are the phase transition curves for solid–liquid, liquid–vapor, and solid–vapor phase changes. Depending on the pressure, water can stay in liquid state up to more than 600 K. The phase diagram of water (Figure 1) reflects the complex behavior of the substance.

On planetary surfaces, temperatures are determined by the solar irradiance, which is a function of the solar luminosity (L) and the distance (D) as L/D^2. Thus, phase conditions for liquid water are dependent on the distance from the central star. Conservative estimates for distances at which water can be stable as a liquid on planetary surfaces in our solar system are between 0.95 and 1.37 AU (Kasting et al., 1993). Thus, of the terrestrial planets, only Mars (apart from Earth) has the potential for transient liquid water on the surface as its orbit is just at the outer edge of the estimate by Kasting et al. (1993). For the Moon, which is inside the distance limits but has no atmosphere, and for Mercury, only ice might be stable on or close to the surface within permanently shadowed regions. The present-day surface of Venus is too hot to support liquid water.

On Mars, with its thin atmosphere, liquid water might be a transient occurrence. If the total pressure of atmospheric gas exceeds 610 Pa, pure liquid water may be stable against boiling, but it is still unstable with respect to evaporation. The boiling point of a liquid is that temperature at which its saturation vapor pressure equals the total external pressure. This total pressure can be supplied in the case of Mars, for example, by CO_2, and does not refer to the partial pressure of water vapor. For example, at 283 K, the saturation vapor pressure of liquid water is $\sim$12 mbar. With these atmospheric conditions, liquid water would be stable against boiling even if the partial pressure of water in the atmosphere is less than the saturation vapor. Since the atmospheres of Mars and Earth are not saturated, liquid water is not stable on the surface of either planet with respect to evaporation. Stability of liquid water on the surface means therefore in general that liquid water, if it could form, would be stable against boiling or freezing but (most likely) unstable to evaporation. For more details, see the discussion by Haberle et al. (2001).

As pointed out by Richardson and Mischna (2005), boiling is only of major importance as a process because it leads to a much more rapid conversion of liquid to vapor than evaporation. Bubble development in the interior of the liquid is allowed and dynamically favored, if the saturation vapor pressure of the liquid exceeds the total atmospheric pressure. It enables the vaporization of the liquid to occur throughout the full volume of liquid. The rough surface created by boiling might also lead to a more effective heat exchange. In general, it can be summarized that liquid water might exist if surface pressures and temperatures fall between the melting point of ice and the boiling point of liquid water.

For an atmosphereless body like Mercury or the Moon, the stability of ice is determined purely by the sublimation rate as discussed recently by Andreas (2007). The standard equation for the sublimation (or evaporation) of ice (or pure water) in vacuum is given by Estermann (1955) as

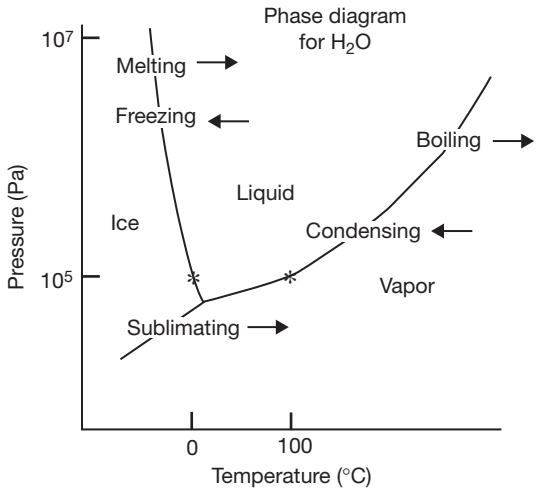


Figure 1 Phase diagram of water.

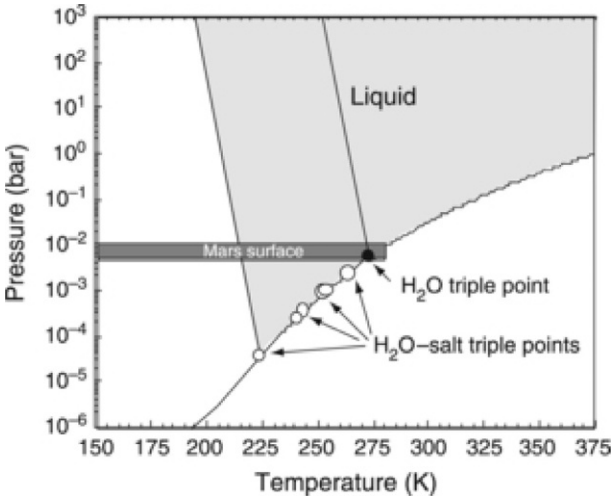


Figure 2 Phase diagram for a H_2O–$CaCl_2$ brine. Reproduced from Bodnar RJ (2001) PTX Phase equilibria in the H_2O–CO_2–salt system at Mars near-surface conditions. *Lunar and Planetary Science* XXXII: 1689 (CD-ROM).

$$S = p_{\mathrm{sat},i}(T)\left[\frac{M_\mathrm{w}}{2\pi RT}\right]^{1/2}$$

where S is a mass flux in kg m^{-2} s^{-1} and $p_{\mathrm{sat},i}(T)$ is the saturation pressure in hPa for an ice surface at temperature T (in Kelvin). M_w is the molecular weight of water and R the universal gas constant. Using the sublimation equation, it is straightforward to derive the temperature-dependent sublimation if the saturation vapor pressure curve is well known especially at low temperatures.

Murphy and Koop (2005) provided recently a new function for the temperature dependence:

$$p_{\mathrm{sat}}(T) = 0.01\exp\left(9.550426 - \frac{5723.265}{T} + 3.3068\ln T - 0.00728332\right)$$

The expression is valid for $T = 110$–273.15 K. $p_{\mathrm{sat}}(T)$ is in hPa, and T in K.

For a planet like Mars, the sublimation can be limited by the saturation of the atmosphere with water vapor. The saturation can be expressed in terms of relative humidity:

$$r_\mathrm{h} = \frac{\rho \cdot R_\mathrm{v} \cdot T}{p_{\mathrm{sat}}(T)}$$

where ρ is the water vapor density, T the temperature, p_{sat} the saturation pressure, and $R_\mathrm{v} = 461.5$ JK^{-1} kg^{-1}.

For Mars, the abundance of water vapor in the atmosphere *pr* is often given in precipitable microns, describing the amount of water if the whole atmospheric column would be condensed on the surface.

Assuming the simple case that water follows a scale height-dependent distribution, this can be converted to water vapor density using

$$\rho(t) = \frac{\rho_{H_2O} \times pr}{H(t) \times 100}$$

where ρ_{H_2O} is the density of water, *pr* the water abundance in precipitable microns, and $H(t)$ the scale height that can vary as a function of Martian seasons. One has to be cautious, however, because the vertical distribution of water in the Martian atmosphere is not readily describable only by a scale height. There are many instances where water does not follow a scale height or where there is an inversion and the number density increases with altitude.

Salinity does not influence the solid–vapor transition temperature, as salts are not incorporated into ice. However, it does influence the solid–liquid transition temperature. Addition of salt to water causes the triple point to migrate to lower temperatures and pressures (see **Figure 2**). Liquid water may exist in the form of salt-saturated solution (also called brines) at temperatures significantly below 273 K. Brines with a high content of $CaCl_2$ remain liquid up to 220–225 K (see, e.g., Brass, 1980; Kuzmin and Zabalueva, 1998). Haberle et al. (2001) had pointed out that while pure water could exist over 29% of the surface of Mars for 37 sols each year, a NaCl eutectic solution could remain liquid over most of the planet at some time of the year and up to 100 sols in some regions. As discussed before, this refers only to transient liquid water, as liquid water is stable nowhere on the Martian surface. The formation of eutectic solutions of $CaCl_2$, with its much lower eutectic temperature, would greatly increase the extent and duration of liquid on Mars. Sears and Chittenden (2005) showed recently that brine formation will increase the stability of water on Mars not only by extending the liquid temperature range but also by considerably decreasing the evaporation rate. At the eutectic temperature of NaCl brine, evaporation rates are ~30 times less than pure water; at the eutectic temperatures of $CaCl_2$, evaporation rates are slightly lower. In longer contact with the atmosphere, the water will still evaporate leaving behind evaporites. Such deposits have been suggested based on the geochemical studies by the Mars Exploration Rovers (MERs) (Christensen et al., 2003; Squyres et al., 2003). For a short time, brines could appear on the Martian surface or nearby even under present-day conditions and would be stable over longer time spans than pure liquid water.

10.11.1.1.2 Adsorption

Hydrogen bonds can form not only between water molecules but also between a water molecule and a silicate surface. The water molecules are typically bound in monolayers to the grain surface. There are a number of theories describing the amount of water that a certain material can hold. Following Jakosky (1985) with σ as the number of molecules bonded, we get

$$\sigma = \frac{Ap(1-\alpha)}{(2\pi MRT)^{1/2}}\tau_0 e^{Q/RT}$$

In this equation, α is the fraction of molecules striking the surface A, which do not stick, and it is assumed that the vibrating water molecules have a Maxwellian energy distribution. Q is the binding energy, R the gas constant, M the molecular weight, and T the temperature.

Defining $\theta = \sigma/\sigma_0$ as the fraction of monolayer of molecules present yields

$$\theta = \frac{kp}{1+kp}$$

where

$$k = \frac{A}{(2\pi MRT)^{1/2}}\frac{\tau}{\sigma_0}(1-\alpha)$$

These two equations form the Langmuir adsorption isotherm (Langmuir, 1915). This is the most simple approach because it limits adsorption to one monolayer. For most real materials, it fails, because they can adsorb more than one monolayer. A more realistic approach was proposed by Brunauer et al. (1938), termed the Brunauer–Emmett–Teller (BET) theory. Using a similar approach as discussed earlier, the isotherm can be expressed as

$$\theta = \frac{kx}{(1-x)(1-x+kx)}$$

where $x = P/P_s$ and P_s is the saturation pressure of the molecule. As adsorbed water molecules behave typically more like a liquid than a solid (quasiliquid), P_s is the saturation pressure with respect to the liquid.

These rather simplified theories of adsorption do not take into account a number of parameters, including grain-size distribution, ambient pressure, and temperature. Applying these theories to Mars, for example, has shown the need for measurements of the adsorption isotherms of real materials and realistic planetary conditions. Fanale and Cannon (1971, 1974) measured the adsorption of H_2O on basalt for $T = 250$–300 K and $P = 2$–2000 Pa. They found that $\sim 10^{-3}$ kg of water per kilogram was adsorbed. To extrapolate their measurement to Mars condition, usually a least squares fit of an empirical isotherm is used (Zent and Quinn, 1997):

$$\rho = \rho_s A_s \frac{\gamma P^{0.51}}{e^{\delta/T}}$$

This equation gives the adsorbed load of water per cubic meter of regolith. Here, ρ_s is the density of the regolith, A_s the specific surface area, $\gamma = 2.0487 \times 10^{-12}$ kg m^{-2} Pa^{-1}, and $\delta = -2679.8$ K.

Using this isotherm and assuming a specific surface area of 17 m^{2}g^{-1} (Ballou et al., 1978), it was predicted that the Martian regolith can hold about 2.1 kg m^{-3} of water and that formation of ground ice would be suppressed by this effective adsorption process (Flasar and Goody, 1976; Toon et al., 1980).

Zent and Quinn (1995) made the first measurements of adsorption of water onto palagonite and montmorillonite at conditions appropriate to Mars. The palagonite data fitted best to a Toth isotherm

$$\theta = K^*P/[1+(K^*P)^{\upsilon}]^{1/\upsilon}$$

and the montmorillonite data to a Freundlich isotherm

$$\theta = (K^*P)^{\upsilon}/[1+(K^*P)]^{\upsilon}$$

with $K^* = K_0 e^{\varepsilon/T}$ for both isotherms. For palagonite, the fitting parameters are $K_0 = 7.54 \times 10^{-9}$, $\varepsilon = 2697.2$, and $\upsilon = 0.4734$, and for montmorillonite, the fitting parameters are $K_0 = 2.207 \times 10^{-9}$, $\varepsilon = 4157.2$, and $\upsilon = 0.6593$.

With these new isotherms acquired under realistic conditions, Zent and Quinn (1995) found that adsorption is about an order of magnitude less efficient than previously thought and the formation of ground ice is not suppressed by adsorption. Zent et al. (2001) extended these studies to clays, namely, smectites. They discovered that they play no significant role for diurnal cycles of water, because the equilibration time is much longer than a Martian day and the abundances of adsorbed water are much smaller than predicted.

In a recent study, Jänchen et al. (2006) showed that microporous minerals (as nontronite, montmorillonite, chabazite, and clinoptilolite) might remain hydrated under present Martian atmospheric conditions and hold up to 2.5–25 wt% of water in their void volumes at a partial water vapor pressure of 0.001 mbar in a temperature range of 333–193 K. These results are in contradiction to the much lower values reported by Zent and Quinn (1997) and Zent et al. (2001). It should be however be noted that the measurements by Jänchen et al. (2006) have been performed at 257–333 K and have only been extrapolated to the Martian temperature range. This extrapolation might be a reason for the disagreement and demonstrates the importance of measurements under realistic Martian conditions.

10.11.2 Observational Evidence

10.11.2.1 Spectroscopy, Morphology, and Radar

10.11.2.1.1 Visual and infrared spectroscopic features of ice and bound water

H_2O is a bent polar molecule and has as such three fundamental vibration modes – a symmetrical stretch mode at 2.738 μm (ν_1), a bending mode at 6.270 μm (ν_2), and an asymmetrical stretch mode at 2.663 μm (ν_3). In the vapor phase, all absorptions predicted by group theory are observed. Phase changes into the liquid and solid phase change the spectral properties of water. In general, hydrogen bonds tend to broaden and intensify the O–H stretching features and shift them to longer wavelengths (Hamilton and Ibers, 1968). In liquid water, the fundamental vibration modes shift to $\nu_1 = 3.106$ μm, $\nu_2 = 6.079$ μm, and $\nu_3 = 2.903$ μm (Clark, 1999).

Water frost and ice are generally bright and spectrally flat in the visual wavelength range. They show various diagnostic

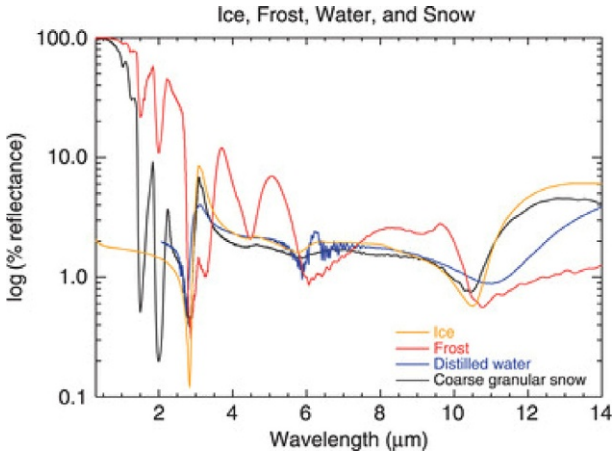


Figure 3 Measured reflectance spectra of different types of water deposits on planetary surface – water ice, frost, distilled water, and coarse granular snow. Data from the ASTER Spectral Library through the courtesy of the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

features in the ultraviolet (UV) and in the near- and mid-infrared (IR) (see **Figure 3**). Water ice has fundamental vibration modes at 3.2 and 6.1 μm and exhibits strong combination overtone bands centered near 1.5 and 2.0 μm. The fact that these bands are so strong makes them prone to saturation. Therefore, the weaker overtones observed at 1.26, 1.04 μm and shorter wavelengths can often provide more useful information (Clark, 1981; Roush et al., 1993). The unsaturated ice bands have distinctive positions, widths, and shapes different from the bands due to molecular water and OH^- absorptions on minerals as will be discussed later.

Many minerals can incorporate H_2O and/or OH^- in their structure. H_2O can also be absorbed on mineral surfaces, located in interlayer sites of smectite clays or trapped as fluid inclusions. All these forms are commonly found in terrestrial rocks as a result of weathering, hydrothermal alteration, or sedimentary precipitation. The near-IR region is ideally suited to search for water- or hydroxyl-bearing minerals (Farmer, 1974; Roush et al., 1993; Salisbury et al., 1991). Dominant in this region is a strong absorption feature around 3 μm. This feature is caused by an overlap of the symmetrical and antisymmetric stretching modes with the first overtone of the strong 6 μm bending mode. The feature is unfortunately not very sensitive to changes in the mineralogy; therefore, it is useful to identify the presence of water but less useful to constrain compositional information on the water-bearing mineral (Roush et al., 1993). Other features in this spectral region, although weaker, are more diagnostic. The region between 2.1 and 2.5 μm shows a group of sharp but weak bands. These are caused because of the stretching fundamental vibrations of cation OH^-, the 1.9 μm overtone of the H–O–H bending mode, and the 1.4 μm OH^- overtones in both H_2O and OH^- (Roush et al., 1993; Salisbury et al., 1991). The fine structure in this region can provide detailed information on the host minerals, because the combining spectral features are due to different physical processes. The 1.9 μm overtone, for example, is caused by adsorbed or bound water in a mineral, like interlayer H_2O in a smectite clay. The presence of the 1.4 μm overtone, combined with the absence of the 1.9 μm absorption feature, indicates the presence of OH^--bearing minerals.

10.11.2.1.2 Gamma-ray spectroscopy and neutron spectroscopy

Gamma-ray spectroscopy and neutron spectroscopy can be used to survey planetary bodies for hydrogen. The abundance of hydrogen is an indirect measurement of the abundance of H_2O or OH^- in the surface layer of a planetary body under certain assumptions. While the detection of hydrogen by these methods is, in contrast to optical spectroscopy, unambiguous, the conversion to water abundance is strongly model-dependent.

Galactic cosmic rays that penetrate the surface layer of a planetary body begin to lose energy immediately through a variety of mechanisms, mainly intranuclear cascades (Bertini, 1969). The secondary particles formed continue to lose energy by a variety of processes. The net result is that neutrons dominate the residual particle population below 100 MeV (Drake et al., 1988; Feldman et al., 1993) and gamma rays produced by the deexcitation of nuclei with characteristic energy spectra. On entering the surrounding material, the generated neutrons are moderated by elastic and inelastic collisions. The resulting spectrum below 10 keV has two distinct energy regimes, the epithermal and the thermal neutrons.

The cross section of hydrogen for the collision with neutrons is very high, and therefore, hydrogen is a very good neutron moderator. The epithermal neutron energy range (0.4–500 keV) is most sensitive for the detection and mapping of hydrogen (Feldman et al., 2002). Simultaneous measurements of the flux of fast, epithermal, and thermal neutrons emitted from a planetary surface allow constraining the amount of hydrogen as well as the vertical distribution in the subsurface. Deriving these constraints requires a multistep modeling approach, including the modeling of the neutron generation, the moderation in the subsurface, and propagation to orbit (Feldman et al., 1993). A parameterization of the neutron energy distribution based on altitude of the observation and depth below the surface has to be derived (Drake et al., 1988; Feldman and Drake, 1986). The model of the depth dependency includes assumptions about the soil composition, which add to the uncertainty of the obtained results.

The detection of hydrogen by gamma rays is based on the characteristic line at 2.2233 MeV generated by the neutron capture reaction (Reedy, 1978). Deriving the abundance and vertical distribution requires again a multistep modeling approach. This modeling again relies on assumptions about the soil composition and structure.

Gamma-ray spectroscopy and neutron spectroscopy allow detecting hydrogen buried in the subsurface. The penetration depths depend on the properties of the regolith and on the distance between detector and surface. Currently, flying instruments such as the gamma-ray spectrometer (GRS) suite on Mars Odyssey can detect hydrogen down to a burial depth of ~1 m (Feldman et al., 2004). This is an advantage over visual and IR spectroscopy, which can detect water only at the surface. The disadvantage of the technique is a model-dependent ambiguity between abundance and burial depths.

10.11.2.1.3 Morphology

The solid surface of planets and moons is shaped by both endogenic and exogenic processes (Greeley, 1985). The resulting surface morphology is due to a combination of these processes. Endogenic processes include tectonic and magmatic activities. The geologic action of water and ice belongs to the exogenic realm and can affect the surface morphology in several ways. Erosion, including transportation, tends to degrade the surface, while deposition of material leads to the formation of new sedimentary bodies. Erosional and depositional features result from the nature and intensity of weathering and erosional processes, which, in turn, are related to climatic and environmental conditions.

Information about the morphology of planetary surfaces comes mainly from spacecraft images, but radar data (e.g., the Magellan data for Venus) and recent highly resolved topographic data from laser altimeters (e.g., Mars Orbiter Laser Altimeter (MOLA) data for Mars) also contributed much to our knowledge of extraterrestrial landforms. The study of landforms is called geomorphology. It is not the intent of this chapter to give an overview of this discipline; instead, the reader is referred to classic textbooks, for example, Twidale (1976), Bloom (1991), Summerfield (1991), and Easterbrook (1999). Typically, processes related to water are classified depending on the environment: fluvial processes are related to running water, lacustrine and marine processes occur in lakes and seas, and glacial processes shape the surface by the action of ice. Water can also alter rocks and therefore change their physical properties, which in turn affects their susceptibility to other geologic processes.

A major problem of the interpretation of remotely sensed data lies in the distinction whether erosion and sedimentation are driven by water or another geologic agent like wind. For example, the sinuous rilles on the Moon were interpreted on the basis of Earth-based telescopic images and Lunar Orbiter photographs as erosional features carved by water by some early researchers (e.g., Urey, 1967). Zimbelman (2001) pointed out that the improvement in spatial resolution alone, as provided by Lunar Orbiter compared to Earth-based telescopic images, could not rule out several of the published hypotheses that are now known to be wrong. Only Apollo images in combination with in situ observations by astronauts together with laboratory investigations of returned samples revealed the volcanic origin of the rilles. Similarly, there is a long debate on Mars whether the huge outflow channels (see **Section 10.11.3.3**) were carved by water or another fluid like liquid CO_2, lava, or wind. It is therefore critical to test several hypotheses before coming to a final conclusion about the nature of morphological features.

The first and for a long time the strongest evidence that Mars has once experienced a warmer and wetter climate came from image data, starting with the Mariner 9 mission of 1971 (e.g., Masursky, 1973) and the Viking orbiter missions, and their morphological interpretation. Since water-related surface features often have a limited spatial extent and the minimum object size to be detected in images depends on the available image resolution (see Zimbelman (2001) for a discussion on image resolution and its role in the photogeologic interpretation), the recent advance of high- and very high-resolution images taken by the MGS, Mars Odyssey, Mars Express, and Mars Reconnaissance Orbiter (*MRO*) missions has greatly enhanced our ability to analyze small-scale morphologies.

Except Mars and Earth, no other terrestrial planet displays morphological features, which can be attributed to flowing water. Surface processes involving flowing water imply an atmosphere, so the lack of it is not surprising on the Moon and Mercury. Venus has an atmosphere, but at least in its current state, it contains no significant amounts of water vapor and it is too warm to allow liquid water or ice.

10.11.2.1.4 Radar

Radar techniques used for planetary exploration are typically ground-based radar, ground-penetrating radar (GPR) on orbiters or rovers, or synthetic aperture radar on orbiters.

Radar at frequencies commonly used for GPR in planetary exploration cannot unambiguously identify water. Dielectric-relaxation responses can uniquely identify water and ice, but these lie at frequencies of >10 GHz and <10 kHz, respectively, well above and below the GPR band, and above 10 GHz, the depth of investigation will be much less than 1 m.

At radar frequencies, rock and ice have representative dielectric constants of 5–10 and 3, respectively. In contrast, liquid water near freezing has a dielectric constant of 87 (Fletcher, 1970; Hasted, 1973; Heggy et al., 2001; Olhoeft, 1981). The dielectric constant of ice exhibits stronger temperature dependence than water, whereas rocks show almost no temperature effect. The strong difference in the dielectric constants for rock and water makes radar a favorable technique for detecting water, especially if low-frequency sounding (<10 kHz) is used that gives access to the dielectric-relaxation response (Grimm, 2002). This is however only true if water is confined in a larger volume – for example, in an aquifer. If water is dispersed in the pore space of a dry regolith, the contrast in the dielectric constant diminishes quickly. Assuming a soil–water mixture with a porosity of 5% yields a dielectric constant of 7–9 indistinguishable from dry rock (Shivola, 1996).

The small dielectric constant of ice allows a good penetration for radar signals. The difference from the typical dielectric values for rocks can result in detectable interfaces between pure ice and dry soil. Mixing of ice and water results in the same problem as described earlier for water–rock mixtures. Even large amounts of ice dispersed within the pore space are indistinguishable from dry rock and can therefore not be detected by GPR.

Ground-based radar observations of planetary surfaces assess the reflective properties of the surface and the near-surface material. Deposits of frozen volatiles on the surface, especially water ice, have a much lower transmission loss than silicate rocks. They produce therefore a higher average radar reflectivity than silicate rocks. This is however not a unique identification because the nature of the surface, especially the spatial-scale roughness of the radar wavelengths, can significantly influence the reflective properties of the surface. The polarization properties of the surface can provide another indication for the existence of ice on or near the surface. Ice-containing surface tends to show a smaller degree of depolarization of the radar signal than dry rock surface. The higher reflectivity and the smaller depolarization effect are due to the details of the multiple scattering at a semitransparent surface (Hapke and Blewett, 1991).

10.11.3 Water on the Surface

If we focus on the surface and near-surface region of terrestrial planets, there are several possible reservoirs for water: in the regolith as ice or adsorbed, on the surface as ice, or as standing or flowing body of water.

There are direct spectroscopic observations of water ice on Mars' polar caps as recently recorded by visible and IR mineralogical mapping spectrometer (OMEGA; Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité) and planetary Fourier spectrometer (PFS) on Mars Express (Bibring et al., 2004b; Hansen et al., 2005) and evidence for its existence today in permanently shadowed regions around the poles of Mercury and the Moon (Butler et al., 1993; Feldman et al., 1998; Harmon and Slade, 1992; Nozette et al., 1996). So far, no direct evidence for liquid water at present time has been found on any terrestrial planet besides Earth. Mars, however, shows evidence that it had in its past large standing and flowing bodies of water. There are also indirect indications for the existence of larger amounts of water in the very early history of Venus.

10.11.3.1 Valley Networks

Valley networks, first detected in Mariner 9 imagery (Masursky, 1973), resemble terrestrial drainage patterns in plan view. More than 1000 individual valley networks were observed on Mars, and their similarity to terrestrial fluvial landforms implies that they were formed by water. They are the most striking morphological evidence that water once flowed across the Martian surface. Nearly all valley networks are located in the heavily cratered highlands, which led to the assumption that they were formed early in Martian history. Later studies used age determinations based on crater statistics and conformed that the majority of valleys were formed at the transition from the Late Noachian to the Early Hesperian epochs (~3.8–3.5 Ga). Although it is commonly accepted that water formed the valley networks, there is a long and ongoing debate about the process (e.g., the relative roles of groundwater versus surface runoff), the source of the water (precipitation as rain or snow versus groundwater or ground ice), and the implications for the climate of early Mars (i.e., 'warm and wet' vs. 'cold and wet'). While the valley networks seem to suggest that liquid water flowed once over the Martian surface for extended periods of time, atmospheric evolution and climate models have problems to explain how a dense, warm atmosphere with a sufficient greenhouse effect could have existed on early Mars (e.g., Fastook et al., 2012; Kasting, 1991; Lammer et al., 2013).

10.11.3.1.1 Morphology and formation processes

Importantly, the term valley network does not describe a river or stream channel. A valley may form by different processes, whereas a river channel is the actual conduit through which liquid water flows. Only relatively few interior river channels on Mars have been observed in high-resolution images (e.g., Irwin et al., 2005b; Malin and Carr, 1999), because they can be easily destroyed by erosion or infilled by sedimentation over long timescales. If their three-dimensional geometry can be resolved, flow models can be applied to constrain the amount of water flowing through these channels per time unit (ranging from a few hundred to a few thousand cubic meters per second, assuming bank-full flow) (Irwin et al., 2005b). Nevertheless, due to their morphological characteristics, most valley networks or systems on Mars are thought to be of fluvial origin.

Valley networks can reach lengths from <5 km to nearly 1000 km and widths from <1 km to nearly 10 km (Mars Channel Working Group, 1983). Some large valley systems, also sometimes termed longitudinal valley systems (examples are Nirgal, Nanedi, Bahram, and Al Qahira Valles), are several hundred kilometers long and several kilometers wide. They exhibit upper reaches with short theater-headed tributaries and a broad, flat-floored main valley with only a few tributaries. On the other hand, small valley systems show a larger morphological variety. Many of them display a dendritic pattern in which tributaries converge downstream. More detailed descriptions about fluvial valley morphology are given in, for example, Baker et al. (1992), Carr (1996), Gulick (2001), and Ansan and Mangold (2013).

Some of the larger or longitudinal valley systems show morphological characteristics (e.g., an abrupt headward termination, prominent structural control, low junction angles, a relatively constant valley width downstream (Goldspiel et al., 1993), and rectangular cross sections with flat floors) that resemble that of terrestrial sapping valleys, which are often situated in layered sedimentary rocks (Howard and Kochel, 1988; Laity and Malin, 1985). Laity and Malin (1985, p. 203) defined sapping process as

> The process that leads to the undermining and collapse of valley head and side walls by weakening or removal of basal support as a result of enhanced weathering and erosion by concentrated fluid flow at a site of seepage.

An origin of Martian longitudinal valleys by groundwater seepage (Goldspiel and Squyres, 2000) would not necessarily require warm and wet surface conditions, and therefore, the problem that atmospheric and climate models do not predict a strong greenhouse effect on early Mars would be less severe. However, groundwater seepage also requires recharge of aquifers by precipitation. Moreover, it has been questioned if sapping alone has enough erosive power to remove the weathered and collapsed sediment in order to continue the retreat of the headwall (Craddock and Howard, 2002; Lamb et al., 2006). Hence, groundwater sapping was probably not the only process involved, and surface runoff may have played an important role, too (Baker et al., 1992; Craddock and Howard, 2002).

The smaller, dendritic valley networks (**Figure 4**) are even less likely to be formed by groundwater seepage alone. Some morphometric properties (e.g., bifurcation ratios, start of branches at high elevations where groundwater emergence is unlikely) support the view that valley network formation required an active water cycle and that surface runoff, either from rain or from snowmelt, was a dominant factor (Ansan and Mangold, 2013). Moreover, whereas studies based on low-resolution data concluded that drainage densities are very low when compared to terrestrial valleys (Carr and Chuang, 1997) and that only some valley systems on volcanoes have comparable drainage densities to fluvial networks on Earth (Gulick and Baker, 1989), measurements based on

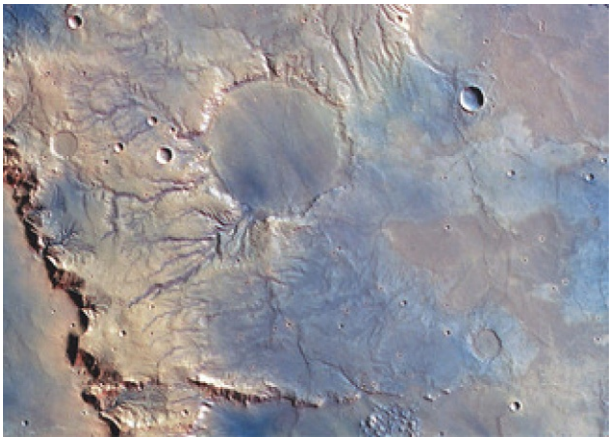


Figure 4 Dendritic valley networks on Mars. The large crater in the center of the image has a diameter of 45 km (false-color HRSC image h00532_0000 at 60° E and 12.5° S). copyright: ESA/DLR/FU Berlin.

higher-resolution imagery data and topographic data have demonstrated that drainage densities on Mars can reach the lower bounds of terrestrial values (Ansan and Mangold, 2006, 2013; Hynek and Phillips, 2003; Hynek et al., 2010; Irwin and Howard, 2002; Stepinski and Stepinski, 2005). Nevertheless, it is still unclear what exactly the climatic conditions were during their formation. Some characteristics of the valley networks, such as their still relatively low drainage densities (Erkeling et al., 2010; Hynek et al., 2010; Luo and Stepinski, 2006; Stepinski and Coradetti, 2004) and other morphometric properties, for example, a paucity of crater rim exit breaches or the shape of longitudinal valley profiles (e.g., Barnhart et al., 2009; Penido et al., 2013), suggest that Martian valley networks are immature as compared to terrestrial drainage patterns and do not indicate an Earth-like climate on Mars that would have been sustained over prolonged periods of time (reviews by Irwin et al. (2008) and Carr (2012)).

10.11.3.1.2 Ages

Most of the valley networks are located in the oldest terrains of the southern highlands, which led to the assumption that they were formed early in Martian history (Carr and Clow, 1981; Pieri, 1976). The age of a larger number of small valley networks was determined by superposition relationships on geologic units. Scott and Dohm (1992) analyzed the transection relationships of about 400 valley networks on geologic units. They derived maximum stratigraphic ages for most valleys between the Noachian and the Early Hesperian. They suggest that possibly 30% of all valleys might be Early Hesperian in age or even younger. Carr (1995) analyzed more than 800 valley networks with the same method and derived that ~92% have Noachian, ~4% Hesperian, and ~4% Amazonian maximum ages. However, these studies suggest that most valley networks were formed early in the Martian history, primarily in the Noachian and Early Hesperian. Carr (1996) pointed out that the cratered highland valley networks are probably mostly Noachian in age and Hesperian-aged ones primarily occur where there is substantial local relief, such as in the Valles Marineris region and along the dichotomy boundary. Most Amazonian-aged valleys are located on volcanoes (Carr, 1995; Gulick and Baker, 1990; Scott and Dohm, 1992). Age determinations by crater counts are complicated due to the small areas of valley floors and subsequent obliteration of craters by infilling processes, such as eolian deposition and mass wasting. The age of some larger valley systems has been determined by crater counts on channel floors (Masursky et al., 1977) and by the so-called buffered crater-counting technique, which takes into account that younger craters hit the valley rims. The problem of small counting areas can be overcome using a so-called buffered crater-counting technique, which takes into account that younger craters hit the valley rims. Applying this method, Fassett and Head (2008a,b) confirmed previous notions that most valley networks are Late Noachian or Early Hesperian in age. Based on global mapping of valley networks, Hoke and Hynek (2009) and Hynek et al. (2010) also concluded that valley networks formed mostly between ~3.8 and 3.6 Ga. This period in the Martian history may have been a peak time of extensive fluvial denudation (Harrison and Grimm, 2005; Howard et al., 2005; Irwin et al., 2005a), since only limited direct observational evidence is left from possible earlier episodes of fluvial degradation (although the degradation state of craters points to significant erosion even before that time).

10.11.3.1.3 Source of water and climate

As pointed out by Craddock and Howard (2002), groundwater flow, recharge, and erosion are closely related to precipitation and surface runoff on Earth. Although sapping is an efficient process in loose sediments, there is no clear evidence from recent studies on Earth that groundwater processes cause significant erosion in resistant rock such as basalt (Lamb et al., 2006). Due to bedrock strength, enormous discharges of water would have been necessary to erode the Martian valley networks (Howard and Kochel, 1988). Estimates of water volumes required for the formation of the valley networks indicate that the aquifers contained about 100–1000 times less water than needed to erode the valleys (Gulick, 2001). A recharge of the drainage areas was therefore necessary (Craddock and Howard, 2002; Goldspiel and Squyres, 2000; Grant, 2000; Gulick, 2001; Howard and Kochel, 1988). The simplest recharge mechanism is by precipitation and subsequent groundwater sapping and/or surface runoff, which would require higher atmospheric temperatures and pressures and point to a warmer and wetter climate on early Mars (e.g., Baker et al., 1991; Carr, 1996; Craddock and Howard, 2002). Phillips et al. (2001) suggested a release of carbon dioxide and water from Tharsis magmas, which may have sustained a warmer climate for the formation of valley networks around or after the end of the Noachian epoch.

Models of valley network formation under dry and cold climatic conditions include the melting of dusty snowpacks by insolation (Clow, 1987), basal melting of snowpacks (Carr and Head, 2003), or melting of snowpacks by internal heat (Zent, 1999). Clifford (1987) proposed an aquifer recharge by basal melting of the south polar cap. However, this model is not consistent with the observation that the elevation of many valley heads is higher than the base of the south polar cap (Carr, 2002). Another suggested recharge model is the global hydrothermal circulation model of Clifford (1993), in which upward vapor flow driven by the

geothermal heat gradient might form an icy permafrost layer. Subsequent melting of permafrost by geothermal heating might cause groundwater sapping (e.g., Squyres and Kasting, 1994). A formation mechanism by snow or ice melt and runoff may specifically apply to valleys of Hesperian and Amazonian age (Fassett et al., 2010; Howard and Moore, 2011).

The genesis of the younger valleys on volcanoes might have been different from valley networks in the highlands. Gulick (2001) suggested that a hydrothermal system associated to magmatic intrusions replenishes itself. Basal melting of a summit snowpack triggered by the heat flux of magmatic intrusions was also proposed for the formation of fluvial valleys on volcanoes (Fassett and Head, 2006).

10.11.3.2 Fluvial and Lacustrine Sedimentary Processes

10.11.3.2.1 The orbiter view

Paleolakes with their morphological characteristics like delta-like deposits, wave-cut terraces, shorelines, and evaporites support the hypothesis of past standing bodies of water on Mars (Cabrol et al., 1996; De Hon, 1992; Forsythe and Zimbelman, 1995; Grin and Cabrol, 1997; Hauber et al., 2009; Newsom et al., 1996; Ori and Mosangini, 1998; Ori et al., 2000; Wharton et al., 1995). More than 200 candidate paleolakes were identified on a global scale in impact craters on Mars, based on fluvial network inflows and morphological lacustrine characteristics (Cabrol and Grin, 1999; Fassett and Head, 2008b; Goudge et al., 2012). In addition to lacustrine sedimentary deposits, a large number of alluvial fans have been identified (Kraal et al., 2008a; Moore and Howard, 2005). Using stratigraphic relationships derived from geologic maps and crater-counting techniques, Cabrol and Grin (2001) estimated most of the paleolakes to be active in the Hesperian, although it has to be noted that the small areas of the deposits may bias the crater-counting results toward younger ages. Nevertheless, a large number of lakes and alluvial fans might have been active even later during the Amazonian, often associated with outflow channels or other short-lived aqueous activity (Cabrol and Grin, 2001; Di Achille et al., 2007; Grant and Wilson, 2011; Hauber et al., 2013). Cabrol and Grin (2001) found little evidence for Noachian lacustrine activity. This apparently contradicts the notion from valley network studies that fluvial activity peaked at the end of the Noachian. Cabrol and Grin (2001) suggested that evidence for it might have been lost due to erosion and major resurfacing episodes later in the Martian history.

Recent image and topographic data allowed detailed analyses of paleolakes and their associated delta and fan deposits and enable reconstructing local hydrologic conditions that help to understand the Martian climatic history. A detailed study of the Tyras Vallis delta deposit in the Xanthe Terra region revealed a complex crater lake history with changing lake levels and water supplies between the Late Noachian and the Hesperian (Di Achille et al., 2006a). Deltaic, coastal, and lacustrine environments occurred during the Hesperian in the Holden crater, followed by glacial processes in Amazonian (Pondrelli et al., 2005). Sedimentary deposits in the Coprates Catena have been interpreted as an alluvial fan (Di Achille et al., 2006b) or delta deposit (Weitz et al., 2006). A particularly well-studied example of deltaic deposits occurs in the Eberswalde crater (**Figure 5**). It shows evidence for channelized and meandering flow indicating persistent flow conditions and formation of intracrater layered sedimentary sequences within fluvial and perhaps lacustrine environments (Malin and Edgett, 2003; Moore et al., 2003). The morphology of this deposit has been interpreted as a fan delta (Bhattacharya et al., 2005; Irwin et al., 2005a; Pondrelli et al., 2008) or alluvial fan (Jerolmack et al., 2004; Malin and Edgett, 2003).

The morphology of deltas and feeder channels has been used to reconstruct the formative processes by modeling and experimental work. These results indicate that the formation of deltas can occur over geologically very short timescales (~years to thousands of years) (Hoke et al., 2014; Kleinhans et al., 2010; Kraal et al., 2008b), eliminating the need to invoke extended periods of a warm and wet climate. However, the presumption of these studies is that, during the epoch of their formation, the deltas were deposited without significant hiatuses or lengthy periods of aridity; if activity on these features was intermittent, the timescale of formation may have been much longer. These results are consistent with the notion of delta formation by local, not global, conditions (Hauber et al., 2013) that may include impacts, volcanism, and transient episodes of warming that were perhaps triggered by orbital or obliquity variations (Mangold et al., 2012). Using laboratory experiments of crater lakes, De Villiers et al. (2013) could show that the water level behavior controls delta morphology: Stepped, retrograding deltas form during water level rise, whereas prograding deltas form during periods of constant water level.

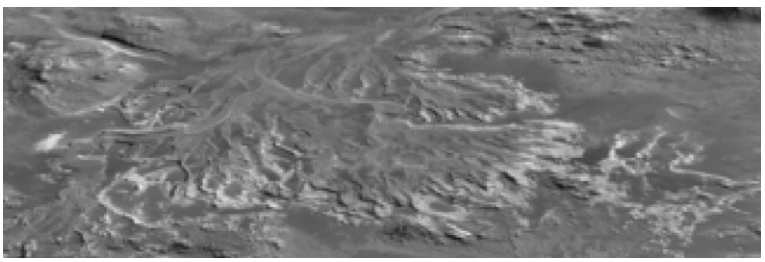


Figure 5 The Eberswalde crater hosts the best-known example of a deltaic structure on Mars. Evidence for channelized and meandering flow indicates persistent flow conditions and formation of intracrater layered sedimentary sequences within fluvial and perhaps lacustrine environments (Malin and Edgett, 2003). The image is a perspective view from the east, and the image width is about 20 km (MOC image mosaic draped over topography derived from HRSC stereo images). HRSC image copyright: ESA/DLR/FU Berlin.

10.11.3.2.2 The view of the rovers

Aqueous processes have also been inferred from morphological observations by the MER rover, Opportunity, in Meridiani Planum (Squyres et al., 2004). The layered rocks consist of sandstones, which show several sedimentary structures like cross-bedding, festoon cross-lamination, resistant fracture fills, and polygonal forms (Grotzinger et al., 2005). Although eolian processes probably dominated the overall environment, these structures, together with geochemical and mineralogical evidence (see **Section 10.11.4.2**), indicate that abundant acidic groundwater was present, that acidic and oxidizing conditions prevailed, and that inundations led to episodic shallow subaqueous flows of liquid water (Squyres et al., 2006). Several terrestrial analogues, ancient and modern, have been proposed to illustrate the paleoenvironment in Meridiani Planum. A summary is given by Benison and Bowen (2006), who favored ephemeral acid saline lakes in SW Australia as the best-known Earth analogue. There, shallow lakes and adjacent groundwater have low pH values (1.5–4) and minerals are precipitated that are also found in Meridiani Planum (gypsum, halite, iron oxides, and jarosite). Flooding, evaporation, desiccation, and winds dominate the environment (Benison and Bowen, 2006). Recently, the Curiosity rover of the Mars Science Laboratory (MSL) mission has investigated in situ fluvial conglomerates as parts of an alluvial fan complex within Gale crater (Williams et al., 2013).

10.11.3.3 Outflow Channels

Outflow channels, first detected in images of the Mariner 9 spacecraft (Masursky, 1973), are the most spectacular examples of Martian surface features formed by the erosive power of water. Their dimensions are much larger than those of other fluvial landforms on Mars as, for example, valley networks (**Section 10.11.3.1**). The largest outflow channels can reach lengths of thousands of kilometers, widths of tens to hundreds of kilometers, and depths of more than 1 km. They are not uniformly spread over the Martian surface, but cluster in a few distinct regions, in particular around the Chryse Planitia basin in the eastern vicinity of the huge magmatectonic center of Tharsis (**Figure 6(b)**: circum-Chryse outflow channels). Other concentrations occur in Elysium Planitia, in the eastern Hellas region, and on the southern margin of Amazonis Planitia. Outflow channels formed episodically in several stages over a significant part of the geologic history of Mars, beginning in the Late Noachian. The main activity seems to have been in the Hesperian period, with lesser stages continuing well into the Amazonian period (see Basilevsky et al., 2006; Table IV of Clifford and Parker, 2001; Greeley and Guest, 1987; Lanz, 2004; Neukum and Hiller, 1981; Scott and Tanaka, 1986; Warner et al., 2010; Werner, 2006). Two outflow channels in the Elysium region, Marte Vallis and the Athabasca Valles, even show evidence for some of the youngest fluvial activity on Mars (Burr, 2003; Burr et al., 2002a,b; Werner et al., 2003).

Outflow channels are important in at least two aspects for the Martian history: first, their enormous sizes strongly indicate that large volumes of liquid water or water ice once flowed across the surface of Mars. Secondly, outflow channels would have been the most obvious pathways for water that filled the northern lowlands to form an ancient Martian ocean (Baker et al., 1991; see also **Section 10.11.3.4**).

10.11.3.3.1 Morphology

Outflow channels start full size at discrete sources. These sources may be fractures (e.g., at Athabasca or Mangala Valles; **Figure 7**: Mangala), but most commonly, they are located in the so-called chaotic terrains (**Figure 6(d)**: Aromatum Chaos (chaotic terrains)). Large blocks of plateau material and intermediate depressions, often arranged in a 'chaotic' geometry, characterize these zones. They seem to have formed by the collapse of plateau material that, where undisturbed, currently forms the surroundings of the chaotic terrains. The most characteristic landforms associated with outflow channels themselves are streamlined or 'teardrop-shaped' islands, which were clearly carved by a fluid medium and indicate the flow direction (**Figure 7**: Mangala). The overall pattern of outflow channels is often anabranching (**Figure 6(b)** and **6(c)**: outflow channels in MOLA). Other surface features typically observed at the floors of outflow channels are bedforms like scour marks, grooves, cataracts, and possible sinkholes ('kolks') (Baker, 1982), testifying that the outflow channels are true river channels (as opposed to valleys; see discussion in Carr, 1996). Outflow channels have only few tributaries and maintain a more or less constant width downstream. More detailed descriptions of the channel morphologies are given by Baker (1982), Baker and Nummedal (1978), and Baker et al. (1992). It was recognized early in the interpretation of spacecraft images that the collective association of these landforms bears a striking similarity to the so-called Channeled Scabland in the northwestern US (Baker, 1978, 1982, 2009; Baker and Kochel, 1979; Baker and Milton, 1974) (**Figure 6(a)**: Landsat). J Harlen Bretz interpreted these giant erosion features in a series of pioneering studies early in the twentieth century as results of the catastrophic drainage of ice-dammed lakes during the latest glaciation in the Pleistocene (e.g., Bretz, 1923). Their significance extends well beyond the Channeled Scabland in that they helped to understand that catastrophic events must not be neglected in the deciphering of the Earth's geologic record ('uniformitarianism' vs. 'catastrophism'; for a comprehensive modern view on this debate, see Baker and Nummedal (1978)). Similar 'spillways' have later been interpreted by means of modern spacecraft data in Siberia (Manych spillway and Turgay spillway; e.g., Baker et al., 1993; Grosswald, 1998).

The cataclysmal flood hypothesis is, however, not without difficulties, as pointed out by Baker et al. (1992). First, under current climatic conditions, water is not stable. Since the floods were probably very short-lived, this does not seem to be a major problem. Second, there are no obvious fluvial deposits like deltas at the termini of the channels. The third problem is that the source areas, which mostly are chaotic terrains, are relatively small when compared to the amount of water that probably carved the outflow channels. This discrepancy might be explained by the headward migration of the source areas as more and more terrains collapse in the subsequent stages of flooding (Baker, 1982).

10.11.3.3.2 Morphometry/discharge

Many attempts have been made to estimate the peak drainage in the outflow channels from the geometry of the

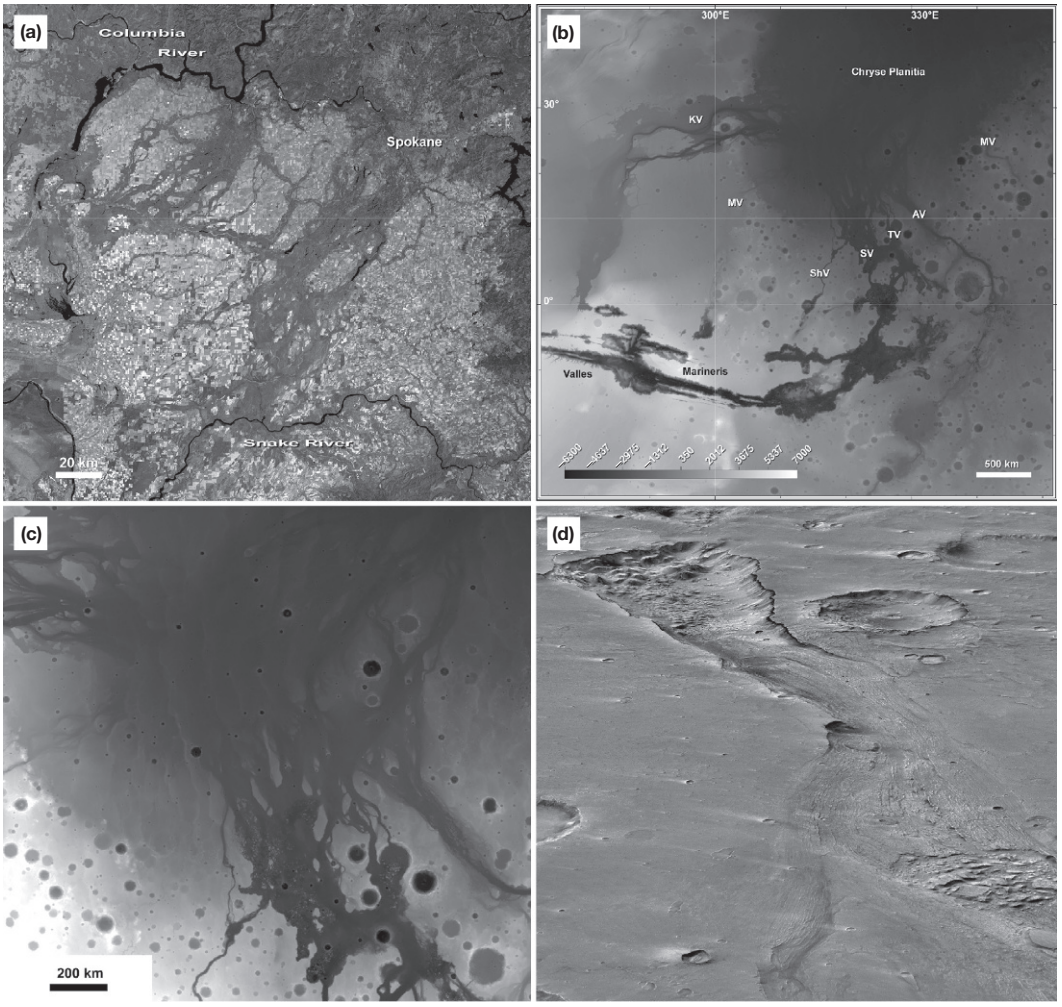


Figure 6 Outflow channels are the most dramatic geomorphic evidence for the flow of liquid water or ice on the Martian surface. They were interpreted as the erosional products of catastrophic floods on the basis of the comparison to terrestrial analogues. (a) The Channeled Scablands in Washington State (the United States) formed when an ice-dammed lake drained in a huge flood toward the Pacific Ocean (Bretz, 1923). The channel floors can be recognized as dark areas and form an anastomosing pattern (Landsat mosaic; north is up). (b) Outflow channel on Mars. The largest channels debouch into the topographic low of the Chryse Planitia basin (KV, Kasei Vallis; MV, Maja Vallis; ShV, Shalbatana Vallis; SV, Simud Vallis; TV, Tiu Vallis; AV, Ares Vallis). The image is a gridded digital elevation model derived from MOLA altimetry data (see scale bar for elevations). (c) Detail from (b), showing the anastomosing nature of the downstream parts of some channels. Note the similarity to the terrestrial analogue shown in (a), but also note the difference in scale. (d) Perspective view of an outflow channel. The source in the left background is a typical chaotic terrain, Aromatum Chaos, and the outflow channel, Ravi Vallis, extends from there to the right foreground, displaying a grooved channel floor (HRSC image mosaic draped over MOLA topography; the diameter of the large crater is about 26 km). copyright: ESA/DLR/FU Berlin.

paleochannels. The first and for a long time the most accepted approach was the application of modified versions of the empirically determined Manning equation (Manning, 1891), which predicts flow velocities in terrestrial channels. In two outstanding papers, Komar (1979, 1980) pioneered studies of this kind and found values in the range of 10^7–$10^8 m^3 s^{-1}$ (for Maja, Ares, and Tiu Valles, assuming a flow depth of 100 m). Even higher values of up to 10^9 $m^3 s^{-1}$ have been reported for Kasei Vallis, the largest outflow channel on Mars (Baker, 2001; Robinson and Tanaka, 1990). Williams et al. (2000) used improved topographic data from the MOLA instrument and reported values between 8×10^4 and 2×10^7 $m^3 s^{-1}$ for channels of the Kasei Vallis system. It has to be emphasized, however, that Komar (1979, 1980) was aware of the shortcomings of this approach (see discussions in Kleinhans, 2005; Wilson et al., 2004). More recently, Wilson et al. (2004) derived a new value for the Manning n coefficient and found that previous analyses based on modified versions of the original Manning equation overestimated the water flow velocities by a factor of up to 2.5. They recommended that their improved value for the Manning n coefficient should be used or, preferably, that a modern form of the Manning equation or the Darcy–Weisbach equation be used for future studies. In an up-to-date review of water flow and sediment transport models, Kleinhans (2005) also recommended the use of the Darcy–Weisbach equation and gave ranges of flow discharges for large outflow channels like Ares or Kasei Vallis in the order of 10^2–$10^3 \times 10^6$ $m^3 s^{-1}$. Discharges within the same order of magnitude were found by Burr et al. (2002a) for the Athabasca Valles ($1–2 \times 10^6$ m^3 s^{-1}), Ghatan et al. (2005) for Mangala Valles ($\sim 5 \times 10^6$ m^3 s^{-1}),

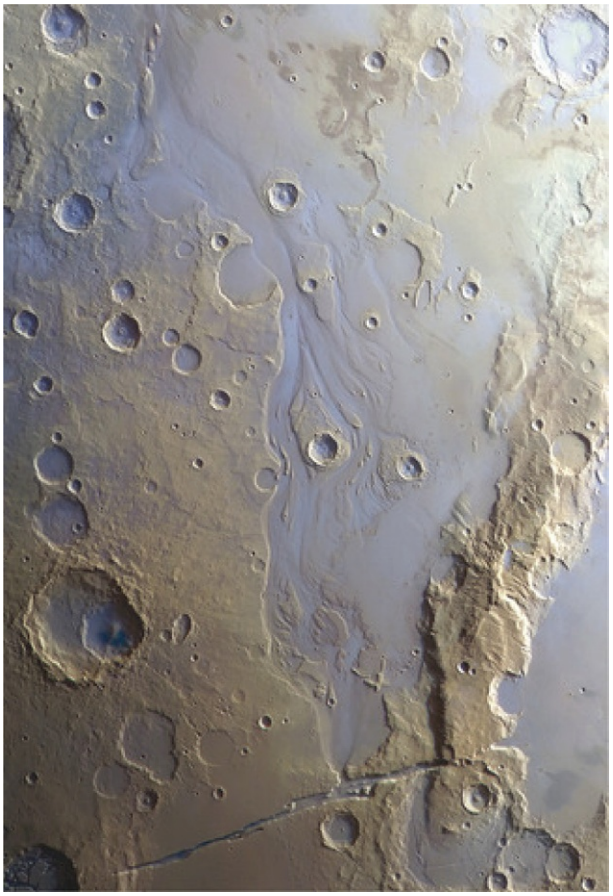


Figure 7 Mangala Vallis, a typical outflow channel on Mars. The channel originates at a tectonic fracture at the bottom center of the image and flowed toward north (top of image), eroding streamlined islands. False-color HRSC image 0644 (image width is about 370 km). copyright: ESA/DLR/FU Berlin.

Coleman (2004) for Ravi Vallis ($1–35 \times 10^6$ m^3 s^{-1}), and Leask et al. (2006a), also for Ravi Vallis (between <1 and 30×10^6 m^3 s^{-1}). Leask et al. (2006a) calculated the duration of the flooding event from the volume of the eroded material, from plausible assumptions on the sediment-carrying capability of water, the minimum total volume of water involved, and the crustal erosion rate. They obtained a duration of 2–10 weeks, comparable with a duration of 1–3 months for the flooding in the Mangala Valles, as reported by Ghatan et al. (2005). The total amount of water required to erode the outflow channels may have corresponded to a global equivalent water layer of 500–1000 m thickness (e.g., Carr, 1986), but this estimate is subject to considerable uncertainties.

10.11.3.3.3 Formation

Many fluid media have been proposed to explain the formation of outflow channels, for example, wind (Cutts and Blasius, 1981), lava (e.g., Carr, 1974; Leverington, 2004, 2011), liquid hydrocarbons (Yung and Pinto, 1978), and liquid CO_2 (Hoffman, 2000a; Milton, 1974; Tanaka et al., 2001); the dissociation of methane hydrate (Max and Clifford, 2001); or liquefaction and/or debris flows (Nummedal, 1978; Nummedal and Prior, 1981; Rodriguez et al., 2006; Tanaka, 1997, 1999) (for a more complete list, see Carr (1996) and references therein). However, the consensus among many researchers is that only liquid water and water ice are physically plausible candidates in the Martian environment (e.g., Baker, 2001; Carr, 1996; Coleman, 2003, 2005), although it has been argued that lava can also be an efficient erosive agent (e.g., Keszthelyi et al., 2014). While the collective assemblage of outflow channel landforms suggests, by analogy with the terrestrial Channeled Scabland, cataclysmal release of liquid water as the formation mechanism, it has to be emphasized that water ice can also explain some of the properties of outflow channels (Lucchitta, 1982, 2001).

How could huge volumes of water be released in short time on the Martian surface? Most researchers think that a pressurized aquifer at depth, covered by an impermeable crustal zone of ground ice (the cryosphere; Clifford, 1993; Lasue et al., 2013a,b), was tapped, and the artesian release of that pressurized water led to the catastrophic drainage of the aquifer, collapse of the surface to form chaotic terrains, and outflow channel formation (e.g., Baker and Kochel, 1979; Baker et al., 1991; Carr, 1979; Masursky et al., 1986; Theilig and Greeley, 1979). Some chaotic terrains are completely enclosed and do not have a surface connection to an outflow channel. On the other hand, there are zones of surface subsidence between some chaotic terrains and outflow channels (e.g., between Ganges Chasma and Shalbatana Vallis) (e.g., Cabrol et al., 1997). Both observations indicate that channelized subsurface flow of water is an important factor in the mobilization of the water masses responsible for outflow channel generation (Carr, 1995). Rodriguez et al. (2003) suggested that extensive cavernous systems, triggered by intrusive magmatism into permafrost, served as pathways for the migration of water that finally emerged into the outflow channels.

Several processes have been invoked to break the thick permafrost zone or cryosphere above a pressurized aquifer, including magmatic intrusions (e.g., Greeley and Spudis, 1981; Mars Channel Working Group, 1983; Masursky et al., 1977, 1986; Mouginis-Mark et al., 1984; Rodriguez et al., 2003), and in particular dikes (e.g., Head et al., 2003b) and sills (Leask et al., 2006b), and tectonism (e.g., Cabrol et al., 1997). Manga (2004) showed for the case of the recent flooding at Cerberus Fossae that large subsurface aquifers with permeabilities comparable to those of young basaltic aquifers on Earth can discharge enough water to account for the estimated 10^6 m^3 s^{-1} (Burr et al., 2002a), without the need for unusual physics or material properties. An alternative way to release water in large amounts might have been the catastrophic drainage of ice-covered lakes in the Valles Marineris area or in the chaotic terrains (e.g., McCauley, 1978; Robinson and Tanaka, 1990). Outflow channels often show evidence for multiple stages of flooding, which requires recharging of the aquifer(s). This could have taken place by basal melting of the topographically high-lying south polar cap (Clifford and Parker, 2001), supplying a globally connected aquifer with water, or by precipitation on the elevated Tharsis bulge in times of higher obliquity (Harrison and Grimm, 2004).

10.11.3.4 Oceans

The Earth is topographically divided in two main regions on a global scale, the high continents and the low-lying,

water-covered ocean floors. Mars is also topographically divided in two main regions on a global scale, the southern highlands and the northern lowlands. The lowlands north of the dichotomy boundary would represent the largest drainage basin on Mars (Banerdt and Vidal, 2001; Smith et al., 2001), and most of the largest erosional features on Mars, the outflow channels (see **Section 10.11.3.3**), debouch into them. Not surprisingly, perhaps, a variety of studies investigated the possibility of an ancient large body of standing water in the lowlands.

10.11.3.4.1 Models involving oceans

Oceans as water reservoirs from which the atmosphere could be replenished constitute major elements in some models of the evolution of the Martian hydrology. Perhaps, one of the most influential was a conceptual scheme outlined by Baker et al. (1991). Overall, this model suggests that the climate since the end of the Noachian was mainly cold and dry, in agreement with the very low erosion rates since then (e.g., Golombek and Bridges, 2000; Golombek et al., 2006). However, short episodes of enhanced hydrologic activity were triggered by volcanism in Tharsis and subsequent cataclysmal outbursts of water that was transported down the outflow channels and finally ponded in the lowlands as oceans, called 'Oceanus Borealis.' The release of the greenhouse gases, H_2O and CO_2, would have created a short-lived denser atmosphere and an active and dynamic hydrologic cycle, which involves precipitation and channel formation (e.g., on areas of high elevations like Alba Patera). In colder regions, glacial landforms would result from snowfall and glaciation. When the released water finally was trapped as ground ice or water, the hydrologic activity would have stopped. The short phases of enhanced hydrologic activities would also have been responsible for the formation of a variety of geomorphic features on Mars, which were otherwise hard to explain. Among these were possible glacial features in the southern hemisphere (Kargel and Strom, 1992) and young, Amazonian-aged valley networks on some volcanoes (Gulick and Baker, 1989, 1990). It has to be noted, however, that recently other possibilities to provide the water for the formation of these young valley networks have been investigated, like the melting of snow (e.g., Fassett and Head, 2006) that was deposited during periods of higher obliquity and enhanced precipitation in lower latitudes.

Most studies proposing an ocean explained it as the consequence of water release through the giant outflow channels. Therefore, its age was commonly assumed to be roughly coeval with outflow channel activity, that is, Hesperian (~3.7–3.3 Ga) or even Early to Middle Amazonian (~3.3–0.6 Ga) in age. In contrast, a recent theory of the hydrologic evolution of Mars postulates the existence of an ocean in the first billion years of the planet's history: Clifford and Parker (2001) investigated the hydraulic conditions that were necessary to explain the outflow channel formation in the Late Hesperian by the catastrophic release of groundwater from confined aquifers. They found that a several-kilometer-thick cryosphere would have been required to support an elevated groundwater table at or above the source region of the outflow channels. In the Noachian, when the climate might have been warmer and the global heat flux was several times higher, the cryosphere would have been too thin to prevent the water from accumulating in the regions of the lowest geopotential. According to Clifford and Parker (2001), the inevitable consequence would have been the formation of a primordial ocean in the Noachian, which might have covered a third of the planet's surface.

10.11.3.4.2 Assessment of geomorphic evidence

One of the first studies that speculated about the existence of an ocean was by Jöns (1984, 1985). He described possible sediments and mudflows (Jöns, 1984) and concluded that the northern lowlands might have hosted an ancient mud ocean fed by mass-wasting processes (Jöns, 1985, 1986). At about the same time, McGill (1985) and Lucchitta et al. (1986) mapped polygonal patterns in the northern lowlands and found a spatial association with the terminations of the big outflow channels. Lucchitta et al. (1986) suggested that the polygons developed in sedimentary layers and concluded that these might have been deposited by outflow channel discharge in an ancient ocean, which could have been covered by ice. Later, but still on the basis of Viking orbiter images with a resolution of typically 60–100 m per pixel, Parker (1989) and Parker et al. (1993) made an attempt to map shorelines around the northern lowlands. They suggested two possible sets of shorelines, an outer set called 'contact 1' and an inner set ('contact 2'). Based on their mapping and a discussion on possibly involved water volumes, Parker et al. (1993) proposed that the northern lowlands have been occupied by an ocean fed by outflow channel activity, the shorelines representing highstands of the water level. For a long time, the shoreline maps by Parker (1989) and Parker et al. (1993) were the foundation of most studies speculating about northern Martian oceans. Later, Edgett and Parker (1997), Clifford and Parker (2001), and Webb (2004) reported on additional shorelines, some of them requiring significant volumes of water to fill the outlined depressions.

In these studies, the outflow channels would have been the pathways for the water filling the hypothesized ancient oceans. Therefore, the plausibility of the ocean hypothesis depends critically on the relationship between the outflow channels and the northern lowlands as the natural sink for their water.

The morphological expression of six major outflow channels disappears in the downstream direction at about the same elevation around Chryse Planitia. Ivanov and Head (2001) noted that this fact might suggest a common base level and would be in agreement with large lakes or oceans in the lowlands in the Hesperian or in the Early Amazonian, the main periods of outflow channel activity (see **Section 10.11.3.3**). Similarly, there are no deltaic deposits observed below one of the shorelines proposed by Parker et al. (1993), which would be consistent with a global-scale base level (i.e., an ocean) (Di Achille and Hynek, 2010). The Vastitas Borealis Formation (VBF), which covers Lower Hesperian ridged plains in the lowlands, is interpreted to be an eroded residue of ice-rich sedimentary deposits, which were emplaced by the floods carving the outflow channels (e.g., Head et al., 2002; Kreslavsky and Head, 2002; Tanaka et al., 2001). The very similar ages of the outflow channels and the VBF support this idea. Further support comes from the approximately identical volumes of the material eroded by the outflow channels and the VBF (Carr and Head, 2003). The dielectric properties of the uppermost crust in the northern lowlands are consistent with

ice-rich material or low-density sediments, which have been interpreted as the remnants of a Late Hesperian-aged (~3 Ga) ocean (Mouginot et al., 2012) fed by the outflow channels.

Polygonal terrain in Utopia, previously suggested to be of sedimentary origin (Lucchitta et al., 1986), was interpreted to be a tectonic consequence of flexural rebound of the crust after removing the load of a former ocean (Hiesinger and Head, 2000). Boyce et al. (2005) examined the morphometry of impact craters in the northern lowlands with MOLA data. A population of Hesperian-aged craters with a low depth-to-diameter relationship seems to be filled by material of the VBF. Boyce et al. (2005) concluded that the crater filling is the result of burial from an ice-rich mantle. They used the extent of this crater population as an indicator of the areal extent of the VBF and the possible limits of an ancient ocean in the Late Hesperian or Early Amazonian.

10.11.3.4.3 Tests of ocean hypotheses

More recent data from the MGS mission (Albee et al., 2001) allowed to test the ocean hypothesis. MOLA data showed that the northern lowlands were extremely smooth and level at scales of a few hundred meters to a few kilometers (Aharonson et al., 1998; Kreslavsky and Head, 1999; Smith et al., 1999). Indeed, they can be compared to the smoothest terrestrial surfaces like the abyssal plains (Smith et al., 1999). Head et al. (1998) and Head et al. (1999) used MOLA data to compare the smoothness of the surface within and outside of the mapped shorelines of Parker et al. (1989). The surface is rougher at all scales from several hundred meters to tens of kilometers outside of contact 1 than between contacts 1 and 2, and similarly, the surface within contact 2 is smoother at all scales than outside contact 2. They also sampled the elevations along the shorelines and found that elevations of contact 1 span almost 11 km, a value that is clearly inconsistent with any plausible equipotential surface in the past. Contact 2 still has an elevation range of ~4.7 km, but the major deviations from an equipotential surface occur in Arabia and Elysium (>1 km downward deflection of the mapped features from an equipotential surface) and in Tharsis (>2 km upward deflection), both being regions where vertical tectonic movements might have occurred after the formation of the shorelines. Head et al. (1999) concluded that MOLA data are consistent with hypotheses of oceans in the Martian past. However, several objections can be raised. First, it seems questionable whether postshoreline downward flexure in Arabia or uplift in Tharsis could really have reached these values in the order of kilometers. Although Clifford and Parker (2001) admitted that errors in the original, Viking-based identification of shorelines by Parker et al. (1989) might have contributed to the wide range of elevations, significant tectonic forces would have been required after outflow channel activity and shoreline formation in the Late Hesperian or even later. Second, detailed analyses of the suggested shoreline features have not shown morphological evidence in very high-resolution MOC images. Malin and Edgett (1999) investigated MOC images crossing the proposed shorelines at several locations and found no evidence for coastal landforms. This result was challenged by Clifford and Parker (2001) who insisted that some of the features shown by Malin and Edgett (1999) against shorelines are indeed evidence in favor of shorelines. However, Carr and Head (2003) in their assessment of the observational evidence for oceans also could not confirm the existence of coastal landforms. While most studies tried to find geomorphic evidence in the form of wave-cut erosional landforms or interior sedimentary features, Ghatan and Zimbelman (2006) searched the shorelines mapped by Parker et al. (1993) and Webb (2004) for coastal constructional landforms but found almost none. Based on this paucity of geomorphic evidence for shorelines, Ghatan and Zimbelman (2006) did not exclude the possibility of a northern ocean, but put constraints on its nature. For example, the water flowing down the outflow channels could have ponded in smaller lakes and not in a global-scale ocean. Such ponds and lakes had already been mapped by Scott et al. (1995) on the basis of Viking orbiter images, but a test of their existence, for example, by examining their potential shorelines using higher-resolution post-Viking data, has not been performed yet. It has to be noted, however, that the detection of shorelines in satellite imagery is difficult even on Earth and that erosion over possibly billions of years would have obscured any such features. Moreover, an ocean would likely freeze very rapidly (Baker, 2001; Carr, 1983; Kreslavsky and Head, 2002) before any significant coastal erosion by wind-generated waves could start.

If there were major bodies of standing water on ancient Mars, it would be expected that sedimentation of carbon dioxide derived from the atmosphere would have led to carbonate formation (e.g., Catling, 1999). Although minor amounts of carbonates have been detected from orbital and in situ measurements (Ehlmann et al., 2008; Morris et al., 2010) and some carbonate-bearing phases are present in the Martian meteorite ALH84001 (Borg et al., 1999), the sparse mineralogical evidence for carbonates has always been a major argument against long-lived oceans and lakes on Mars. The thermal emission system (TES) on board of Global Surveyor did not show spectral signatures of carbonates (Christensen et al., 2001a; Stockstill et al., 2005). Only small amounts are distributed in the Martian dust (Bandfield et al., 2003). The OMEGA imaging spectrometer confirmed the results, reporting no carbonates above the instrument's detection limit (Bibring et al., 2005). Moreover, the identification of olivine, a mineral that is easily destroyed by chemical weathering, might argue against long-lived reservoirs of liquid water at the surface of Mars. It has to be noted, however, that the lack of carbonates does not necessarily exclude the possibility of oceans on ancient Mars. Even minor amounts of sulfur dioxide would produce acidic waters and prevent the formation of carbonates, instead leading to sulfate deposition (Fairén et al., 2004; Lane et al., 2004). Calcium sulfates have indeed been found in orbital data by Gendrin et al. (2005) and Langevin et al. (2005a) and in situ by the MER or Opportunity in Meridiani Planum (Farrand et al., 2013) and by the MSL rover Curiosity in Gale crater (Vaniman et al., 2014).

While the ocean hypothesis is attractive to explain several observations related to the hydrologic evolution of Mars, the existence of an ocean is yet to be proven. Meanwhile, many key questions are still open (summarized by Ghatan and Zimbelman (2006)): Was there a Noachian ocean as proposed by Clifford and Parker (2001)? Was there a genetic relationship between outflow channel discharge and water ponding as an ocean in the northern lowlands during the Hesperian? Was ocean formation a repeated phenomenon (Baker et al., 1991)? Was an ocean liquid, or did it freeze rapidly and then

sublimate? It seems questionable if new remote-sensing data can answer these questions. So far, the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) on Mars Express did not reveal the existence of massive ice deposits beneath the surface of the northern plains, a result that, if confirmed, argues against the evolutionary model outlined by Clifford and Parker (2001). Given the old age of the hypothesized oceans and the susceptibility of coastal landforms to erosion, the search for coastlines with imaging and topographic data is considered to be difficult. Even if the images of the High Resolution Imaging Science Experiment (HiRISE) camera with a lateral resolution of 30 cm per pixel (McEwen et al., 2007a,b) do not show any unequivocal evidence, this result does not disprove an ocean. Perhaps, the best chances to reach more definitive results will involve lander experiments (e.g., with GPR experiments) and sample return missions.

10.11.3.5 Glacial, Permafrost, and Ground-Ice Features

10.11.3.5.1 Glacial landforms

Glacial processes on Earth produce a wide variety of landforms at different scales (e.g., Benn and Evans, 1998; Bennett and Glasser, 1996; Embleton and King, 1975; Evans, 2003; Hambrey, 1994; Menzies, 1995; Sugden and John, 1976). Examples of glacial erosional features include striated rock surfaces at small scales, roches moutonées at intermediate scales, and cirques, troughs, and fjords at large scales. Well-known glacial depositional features are, for example, moraines and till deposits. As part of a glacial system, not only the processes at the glacier surface boundary but also those landforms produced by fluvial discharge from glacier melting are of importance.

The surface of Mars has probably been a cold environment over most of the planet's history (e.g., Bibring et al., 2006; Carr, 1996; Solomon et al., 2005). This and the fact that water, although in unknown quantity, is assumed to have been present at the surface throughout most of this time raise the possibility that glaciers have formed on the Martian surface and that glacial surface features are still observable today.

Mariner and Viking images showed landforms that were indeed interpreted as glacial in origin, but the number of studies was limited. Williams (1978) and Lucchitta (1981) described landforms resembling recessional moraines at the western and northwestern flanks of the large Tharsis shield volcanoes, which are also shown as possible glacial deposits on the global-scale map by Scott and Tanaka (1986). The morphology of some outflow channels is also indicative of glacial action (Lucchitta, 1982, 2001) (see **Section 10.11.3.5**), and both flood and glacial processes may have simultaneously or sequentially shaped them (Costard and Baker, 2001).

Intriguing evidence for glacial processes reported in the Viking era includes esker-like and moraine-like landforms in both hemispheres (Kargel and Strom, 1992). Kargel et al. (1995) reported an assemblage of features in the northern plains that resemble terrestrial glacial landforms, including moraines, tunnel valleys, and possible eskers.

These landforms were recognized to be relatively young (Kargel and Strom, 1992). Since glaciers need to be recharged by precipitation and a water cycle, their formation late in the Martian history would require a major climate change on Mars relatively recently. The scenario of Baker et al. (1991) accounted for this problem in postulating the episodic formation of oceans by outflow channel activity (see **Section 10.11.3.3**), which then triggered short-lived episodes of a warmer climate with precipitation. It has to be noted, however, that the interpretation of these features as being glacial was partly met with skepticism (Carr, 1996), since the low amount of erosion in the Amazonian (e.g., Golombek and Bridges, 2000; Golombek et al., 2006) seems to contradict major climate excursions late in the Martian history.

Images with higher spatial resolution (MOC, THEMIS-VIS, and HRSC) revealed more details of the moraine-like features. Head and Marchant (2003), Shean et al. (2005), and Milkovich et al. (2006) investigated the structures on the northwestern flanks of the Tharsis Montes (Arsia, Pavonis, and Ascraeus Montes) and Olympus Mons (**Figure 8**) and found an assemblage of landforms that they interpreted as the result of cold-based glaciation, in analogy with cold-based glaciers in Antarctica, which are frozen to the ground and do not move by sliding on the ground, but by internal deformation. Bed erosion at cold-based glaciers is limited due to the fact that ice usually does not reach its melting temperature. Hauber et al. (2005) detected possible glacial features at the northwestern flank of Hecates Tholus, where previously no glacial features had been identified. Additional evidence for midlatitude glaciation was described by Head et al. (2005, 2006).

A particular type of ice-related morphology is hypothesized by some authors to exist near the Martian equator in Elysium Planitia. Rafts of darker material seem to have flown on a brighter matrix. In places, the dark material obviously broke into smaller pieces. On the basis of its overall similarity to terrestrial shelf ice, Brakenridge (1993) suggested that quite recent surface water releases filled a topographic depression, that ice floes formed almost immediately, and that either a grounded ice or an ice-covered sea still persists. Later, the area was revisited by Murray et al. (2005), who analyzed HRSC images and basically came to the same conclusions. The massive water release required to fill the roughly North Sea-sized

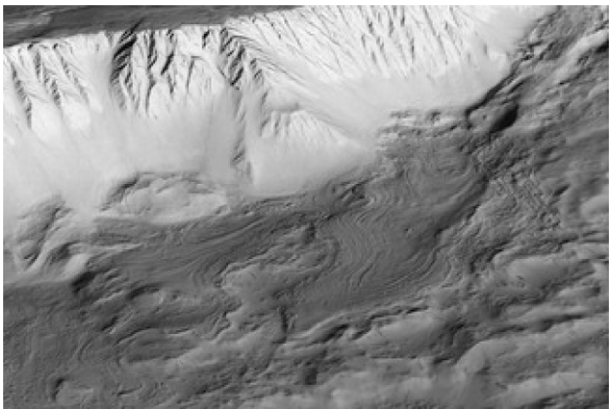


Figure 8 Lobate deposits at the foot of the western scarp of Olympus Mons. Such deposits were interpreted as possible glacial landforms by Lucchitta (1981). Milkovich et al. (2006) ascribed their origin to episodes of increased obliquity during the Late Amazonian period of Mars. Perspective view derived from HRSC stereo images (orbit 2091). copyright: ESA/DLR/FU Berlin.

basin would imply a large underground reservoir. It seems possible that water was transported through the Athabasca Valles, a young outflow channel, toward Elysium Planitia. It must be noted, however, that the pack ice theory for the platy landforms in Elysium Planitia is highly contentious. An origin by flowing lava, which cools at the surface and builds a solidified upper crust that can break in a way similar to pack ice, also seems to be possible (Keszthelyi et al., 2004), and recent very high-resolution images revealed textural details that resemble lava coils only found in terrestrial lava lakes (Ryan and Christensen, 2012).

10.11.3.5.2 Permafrost and ground-ice landforms

Permafrost on Earth is defined as soil or rock that remains below 0 °C for two or more consecutive years, regardless of the water content. Following this definition, Mars may well be regarded as a permafrost planet. Beneath a thin surficial layer, which is affected by the diurnal temperature wave, all of its upper crust can be considered to have temperatures below 0 °C. Ground ice on Earth is ice in frozen ground. Both permafrost and ground ice belong to the realm of the periglacial zone. The term periglacial was first introduced by Lozinski (1909), and in 1910, the concept of a 'periglacial zone' was established to describe climatic and geomorphic phenomena occurring peripheral to the Pleistocene ice sheets (e.g., Huggett, 2003). The term 'periglacial' is nowadays used in a wider sense, and periglacial areas are not restricted anymore to any location or time; it is generally qualitatively used to define areas of cold-climate environments where permafrost might be present although it is not a prerequisite (e.g., Washburn, 1979). Permafrost and ground ice produce a wide variety of landforms on Earth (e.g., Embleton and King, 1975; French, 1996; Jahn, 1975; Washburn, 1979; Yershov, 1998).

In the Viking era, permafrost and ground ice-related landforms on Mars were reported more frequently than glacial ones. Early reports noted the morphological similarity of many Martian landforms to terrestrial analogues of the periglacial zone (Carr and Schaber, 1977; Lucchitta, 1981; Rossbacher and Judson, 1981). The advent of higher-resolution images and detailed global topography obtained through the MGS spacecraft (Albee et al., 1998) spurred a large number of studies analyzing landforms on Mars reminiscent of terrestrial periglacial landforms.

10.11.3.5.2.1 Lobate debris aprons, lineated valley fill, and concentric crater fill

Mass wasting phenomena that display similarities to terrestrial rock glaciers were first described in detail by Squyres (1978) and Squyres (1979). He subdivided them on the basis of their morphology into lobate debris aprons (**Figure 9**), lineated valley fill, and concentric crater fill. Lobate debris aprons are characterized by a convex-upward topography and a steep and distinct flow front, strongly suggesting plastic deformation and flow. They can also show lineations on their surfaces, which can be both parallel and perpendicular to the flow direction. Where lobate debris aprons are confined by valley walls, their surfaces are deformed into contractional lineations parallel to the confining walls. Material with this specific surface texture is called lineated valley fill (**Figure 10(a)** and **10(b)**). Lobate debris aprons and lineated valley fill are best developed

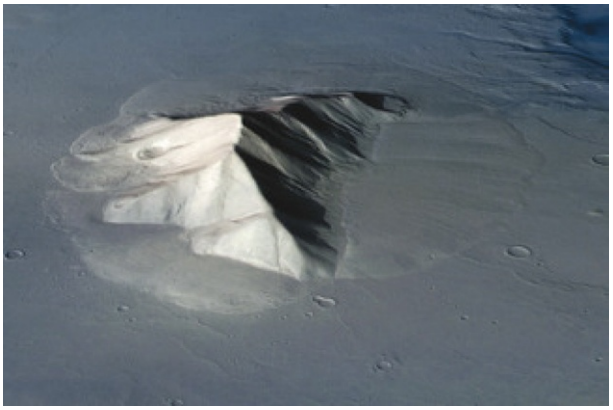


Figure 9 Lobate debris apron surrounding a mountain in the southern midlatitudes (~41° S, 103° E) on Mars. The morphology of the apron suggests viscous flow. Such deposits have been compared to terrestrial rock glaciers. Perspective false-color view of HRSC image 0451. The diameter of the debris apron is about 50 km (north is down, toward the viewer). copyright: ESA/DLR/FU Berlin.

where mesas and valleys with steep walls are found, in particular at the steep scarps of the dichotomy boundary. Concentric crater fill is found in the interior of many craters, where ridges and troughs show a concentric pattern that can also show lobate flow fronts toward the crater center. All these types of landforms show evidence for plastic deformation and resemble terrestrial rock glaciers (Lucchitta, 1984; Squyres, 1978, 1979, 1989), which are mixtures of rock and ice (Barsch, 1996; Wahrhaftig and Cox, 1959). Whalley and Azizi (2003) gave a comprehensive review on the terminology and problems related to the identification of rock glaciers on Mars.

Several hypotheses try to explain the origin of the ice in these debris flows. Water ice could have formed by direct condensation of ice from the atmosphere (Squyres, 1978) or by snow precipitation (Squyres, 1989). It could also accumulate by water vapor diffusion down into the regolith and subsequent condensation (Mellon and Jakosky, 1995). Finally, groundwater may seep into debris and create interstitial ice (Lucchitta, 1984; Mangold and Allemand, 2001; Squyres, 1989). The elastic particles in the lobate debris aprons might come from rock falls that accumulated at the base of scarps (Colaprete and Jakosky, 1998; Squyres, 1978) or, alternatively, from landslides (Lucchitta, 1984; Mangold and Allemand, 2001).

It has been demonstrated by Squyres (1978) on the basis of photoclinometry and by Mangold and Allemand (2001) and Li et al. (2005) on the basis of MOLA topographic profiles that the cross-sectional shape of lobate debris aprons can be approximated by the flow law of polycrystalline ice (Glen, 1955) and the flow relation of ice (Paterson, 1994; Vialov, 1958). Colaprete and Jakosky (1998) modeled flow of ice under Martian conditions and found that flow rates are very low. To create lobate debris aprons of the observed size, they showed that temperatures 20 to 40 K higher than present average midlatitude temperatures (~210 K), ice contents exceeding 80%, and net accumulation rates of $\geq$1 cm per year are required.

Lobate debris aprons, lineated valley fill, and concentric crater fill are young landforms. Crater counts yield low crater densities, and absolute ages of $<$100 Ma have been derived

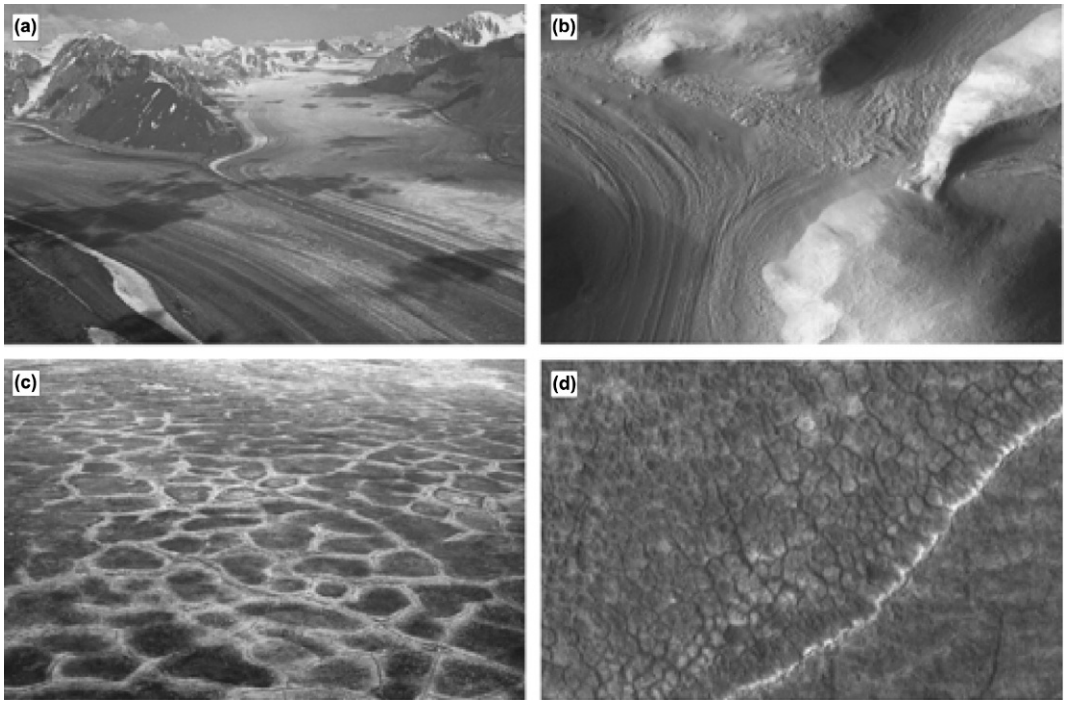


Figure 10 Comparisons between cold-climate landforms on Earth (left) and Mars (right). (a) Glaciers fusing together in the Yukon region (Canada; image courtesy Marli Miller, University of Oregon; image source: Earth Science World Image Bank). (b) Convergent flow features on Mars, probably the result of the flow of rock glacier-like material (MOC image SP245006). (c) Ice-wedge polygons in terrestrial permafrost terrain (Yukon, Canada; image courtesy Agriculture and Agri-Food Canada). (d) Patterned ground (polygons) on Mars. The diameter of the smaller polygons is <10 m and comparable to the size of ice-wedge polygons on Earth (HiRISE image TRA_000856_2265). Image credit NASA/JPL/University of Arizona.

(e.g., Berman et al., 2003; Head et al., 2005; Li et al., 2005; Mangold, 2003; Squyres, 1978). However, morphometric evidence for former lobate debris that were flooded by lava flows or debris flows (e.g., lahars) more than 1 Ga ago suggests that much older rock glaciers existed equatorward of 30°, indicating a paleoclimate allowing ground ice to be stable at that latitudes (Hauber et al., 2008). Indeed, a recent study concluded that widespread glaciation on the northern midlatitudes spanned a range of at least 600 Ma and ended roughly 100 Ma (Fassett et al., 2014).

10.11.3.5.2.2 Terrain softening

At many locations, the Martian topography exhibits a rounded and subdued appearance when compared to undegraded terrain. The effect can be observed in particular at impact craters, which have rounded rims and convex-upward inner wall slopes. This phenomenon was termed 'terrain softening' by Squyres and Carr (1986) and mapped on a global scale. They found that this topographic softening affects terrain in the 30°–60° latitude belts. It affects all sorts of topographic features (Squyres, 1989) and was attributed to the viscous relaxation of the topography due to the creep of a deformable and ice-rich regolith. While the mapping of Squyres and Carr (1986) was done qualitatively on the basis of Viking orbiter images, Jankowski and Squyres (1992) quantified the effect of terrain softening by examining the resulting topography by means of photoclinometry. Theoretical modeling suggests that the thickness of the deformable layer is limited to about 1 km (Jankowski and Squyres, 1992). The effect of viscous creep facilitated by ground ice was also suggested to explain the fact that rocky debris slopes are well below the angle of repose at higher latitudes (Perron et al., 2003).

It has to be noted, however, that mantling deposits can also subdue the topographic signature of the surface. Soderblom et al. (1973) observed a mantling deposit with a similar extent than the terrain softening mapped by Squyres and Carr (1986). They attributed it to dust deposits from the polar regions. The effects of terrain softening and mantling can be discriminated because mantling processes tend to leave sharp topographic crests like crater rims unaffected, whereas terrain softening affects them strongly. In addition, mantling deposits remove small topographic features, while relaxation processes leave them due to insufficient stresses (Squyres et al., 1992).

MOLA data revealed that the global surface roughness displays a latitudinal trend (Kreslavsky and Head, 2000): at latitudes >30–50° in both hemispheres, Mars seems to be smoother at short base lengths than at lower latitudes. This observation is in agreement with the mapping of Squyres and Carr (1986). Kreslavsky and Head (2000) suggested that this trend might be the result of a high-latitude mantling deposit that is several meters thick. Independently, it was found that images taken through the MOC show a very particular texture at certain locations. They reveal a dissected surficial layer or mantle, which suggests the degradation of an originally volatile-rich material (Carr, 2001; Malin and Edgett, 2001; Mustard et al., 2001). Poleward of about 70°, the entire surface appears to be mantled, so that the surface looks very smooth and lacks sharp topographic gradients.

Mustard et al. (2001) suggested that the mantle was deposited as airfall dust and cemented by atmospherically derived ice. There are very few fresh impact craters superimposed on this surface, suggesting that it formed and degraded very recently. Global mapping showed that the mantle seems to be degraded in two latitude belts between 30° and 60° in both latitudes (Mustard et al., 2001). These authors explicitly considered the idea of modern Martian ice ages operating at Milankovitch timescales, triggered by changes in insolation driven by variations in Mars' orbital parameters.

On a global scale, Milliken et al. (2003) mapped the occurrence of young gullies (Malin and Edgett, 2000) (see **Section 10.11.3.8**), viscous flow features (e.g., Squyres and Carr, 1986), and the dissected terrain described by Mustard et al. (2001). Milliken et al. (2003) found a latitudinal concentration of these landforms at midlatitudes with a spatial clustering at latitudes of ~40° in both hemispheres. They suggested that this correlation might be caused by the deposition, deformation, and removal of ice-rich materials at midlatitudes in the Amazonian.

10.11.3.5.2.3 Thermokarst

Thermokarst landforms develop as a result from the thawing of ice-rich permafrost or the melting of massive ice. Typical examples are thermokarst lakes, alases (large depression of the ground surface produced by thawing of a large area of very thick and exceedingly ice-rich permafrost), and thermokarst mounds (hummocks remaining after melting of the ice wedges surrounding an ice-wedge polygon) (van Everdingen, 2005). On Mars, irregularly shaped depressions, mainly in the outflow channels (Costard and Baker, 2001; Theilig and Greeley, 1979), and in the northern plains (Bun Tseung and Soare, 2006; Costard and Kargel, 1995; Morgenstern et al., 2007; Séjourné et al., 2011; Ulrich et al., 2010) and southern hemisphere (Lefort et al., 2010; Zanetti et al., 2010) have been interpreted as thermokarst features.

A fluvial-induced mechanism of thermokarst on Mars has been proposed by Costard et al. (1999). It is based on analogy to huge terrestrial outflow channel-like streams in Siberia during snowmelt in spring, where the high water discharge transfers heat from the flow of water to the frozen ground and triggers ground thawing and thermal erosion. Costard et al. (1999) suggested that relatively warm floods on Mars might have enlarged outflow channels by a combination of thermal and mechanical erosion along frozen river banks.

10.11.3.5.2.4 Patterned ground

A variety of landforms on Mars resemble terrestrial patterned ground. On Earth, this category of landforms includes polygons, circles, stripes, and nets (e.g., French, 1996; Washburn, 1979). Polygons, a common surface phenomenon in terrestrial permafrost areas (e.g., Lachenbruch, 1962), are the most frequently observed periglacial-like features on Mars. Viking orbiter images showed several examples; however, most of them had sizes that were an order of magnitude larger than their possible terrestrial analogues (Rossbacher and Judson, 1981). For example, giant polygons in Utopia Planitia with diameters of 20 km are probably too large to have originated by thermal cracking (Pechmann, 1980) and might instead be the product of fracturing of a sedimentary or volcanic layer due to cooling or desiccation shrinkage plus differential compaction over a rough buried surface (McGill, 1986) or the result of tectonic uplift (Hiesinger and Head, 2000).

Possible periglacial landforms were described in more detail on the basis of post-Viking data with a much higher spatial resolution. Small-scale polygons (**Figure 10(c)** and **10(d)**) could now be analyzed morphometrically, and MOC images revealed that their size distribution is often much more like that of terrestrial ice-wedge polygons than it was thought before (Kuzmin and Zabalueva, 2003; Seibert and Kargel, 2001; Yoshikawa, 2003). Recent images with a resolution of less than 1 m, obtained by the HiRISE camera on board the MRO (McEwen et al., 2007a,b), show that the diameters of polygons in Utopia Planitia are in the order of 10 m, which corresponds exactly to the diameters of their terrestrial counterparts. Global-scale mapping using MOC images showed that the polygons are concentrated in two latitudinal belts poleward of 55° (Mangold, 2005) and correlate to the distribution of ground ice detected by the neutron spectrometer aboard Mars Odyssey (Mangold et al., 2004a,b). However, the absence of polygonal fracture patterns in the north polar area and a high abundance of these landforms near the south polar cap suggest formation processes connected to seasonal deposition of CO_2 ice. Seasonal changes in polygon patterns were observed by van Gasselt et al. (2005) on the basis of multitemporal MOC images of a trough on the south polar cap. They ascribed these changes to contraction-crack processes in a thin layer of water ice beneath the seasonal CO_2 ice cover.

10.11.3.5.2.5 Pingos and frost mounds

Pingos on Earth are perennial frost mounds consisting of a core of massive ice, produced primarily by injection of water (Mackay, 1973, 1979; Porsild, 1938; Washburn, 1979). The identification of pingos in Viking orbiter images has been ambiguous due to the relatively low image resolution (Lucchitta, 1981, 1985; Rossbacher and Judson, 1981). Higher-resolution MOC images allow a closer inspection of candidate topographic features, and several authors cite evidence for pingos at diverse locations like Gusev crater (Cabrol et al., 2000), Athabasca Valles (Burr et al., 2005), and the Cerberus Plains (Page and Murray, 2006), all at equatorial latitudes, and in the midlatitude region of Utopia Planitia (Soare et al., 2005).

10.11.3.6 Polar Caps

The inventory and the seasonal change of the water and carbon dioxide reservoir in the polar regions of Mars play an important role in understanding the evolution of the planet and its current state.

Viking data showed that the residual polar caps consist of H_2O ice. The north polar cap loses a seasonal cover of CO_2 ice annually and the residual consists of H_2O ice (Farmer et al., 1976; Kieffer et al., 1976) The south polar cap partly retains its CO_2 cover (Kieffer, 1979; Paige et al., 1990). This survival of the CO_2 ice veneer on the south polar cap was explained with a lower concentration of dust and therefore higher albedo (James et al., 1992). However, it was unknown if the residual CO_2 ice is underlain by H_2O ice (Thomas et al., 1992) The layered deposits in both polar regions have been assumed to

consist of H_2O ice (Cutts, 1973; Herkenhoff and Murray, 1990).

Except for some spatially unresolved 10–50 μm IRIS spectra (Paige et al., 1990) and TES spectra (Kieffer et al., 2000), the spectral characteristics of the residual ice caps had not been published from any mission before Mars Express. The OMEGA (Bibring et al., 2004a) and PFS (Hansen et al., 2005) instruments on Mars Express independently detected and mapped spectral signatures of water ice on the southern polar cap of Mars. As expected, OMEGA identified the 1.5 and 2.0 μm spectral signatures of water ice over extended parts of the polar cap. In the bright region, the water ice is seen as a small admixture to CO_2 ice, whereas at the edge of the bright cap, pure water ice with a dust admixture was seen. Surprisingly, OMEGA detected also extended deposits of water ice in large areas tens of kilometers away from the polar cap. PFS – using a much higher spectral resolution but without contiguous coverage – observed a similar distribution of ice; however, there is a significant disagreement in the amount of water ice (Hansen et al., 2005). While OMEGA measurements indicate a H_2O/CO_2 ratio of 15%, the PFS measurements indicate not more than several tens of parts per million by mass. The disagreement is most likely due to different modeling assumptions. While Bibring et al. (2004a) assumed a molecular mixture of H_2O and CO_2, the analysis by Hansen et al. (2005) is based on a discrete condensation and therefore a heterogeneous mixture. Due to the spectral range used by the instruments, they are inherently limited to probing the upper few millimeters of the surface and allow no conclusion about the bulk composition.

Studies of the evolution of the northern polar cap indicate that observed brightness variations are due to changes in the grain size of the water ice (Langevin et al., 2005b). In early summer, seasonal water-ice frost with grain sizes <100 μm sublimates and larger grains ($\sim$1 mm) of the permanent ice cap dominate the optical near-IR spectral properties of the water ice.

Global neutron mapping of Mars at various spectral ranges obtained by the HEND instrument on MGS allowed to constrain the content of water ice and adsorbed and bound water in the upper 1–2 m of the Martian surface. The data returned by HEND showed huge regions of permafrost present both north of 55° N and south of 60° S. As discussed earlier, the measurement of one energy range of neutron flux alone is not sufficient to derive the water-ice content and distribution. However, combining data from neutrons with different energies and 2.2 MeV flux (Boynton et al., 2002; Feldman et al., 2002; Mitrofanov et al., 2002) indicates that a soil model consisting of a layer with water ice overlain by a layer of drier soil is in best agreement with the available data. The lower ice-rich layer contains between 50 and 75 wt% water, with a higher abundance of water ice in the shallow subsurface at northern polar latitudes than in the south (Mitrofanov et al., 2003). These ground-ice deposits extend over $\sim$15% of the surface of Mars and are therefore a substantial water reservoir.

Previous compositional and stratigraphic interpretations of the polar cap deposits have been based on imaging, spectral, thermal, and topographic measurements. With the MARSIS on Mars Express, it was possible to study the internal structure of the caps for the first time. In an early MARSIS orbit, 1855, the

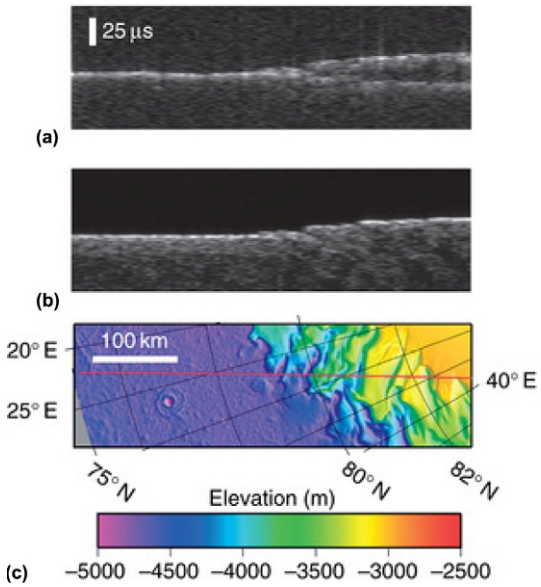


Figure 11 (a) MARSIS data in radargram format for orbit 1855 as it crossed the margin of the NPLD. (b) Simulated MARSIS data if echoes are only from the surface (nadir and off-nadir clutter). (c) MOLA topography along the ground track (red line); elevation is relative to mean planetary radius. MARSIS data at 5 MHz show a split of the strong return into two as the ground track reaches the NPLD (higher terrain to the right). Maximum time delay to the second reflector is 21 μs, equivalent to 1.8 km depth in water ice. Reproduced from Picardi G, Plaut JJ, Biccari D, et al. (2005) Radar soundings of the subsurface of Mars. *Science* 310: 1925–1928.

northern polar layered deposits were briefly observed in the longitude range from 10° to 40° E (Picardi et al., 2005). The radargrams show the surface reflection splitting into a pair of strong reflectors as the ground track passes from the northern plains onto the layered deposits (**Figure 11**). An analysis of the signal taking into account the effect of surface topography as a source for clutter indicates that the second reflector is below the surface. The derived dielectric properties for the overlaying material indicate a fairly pure water ice with $\sim$2% impurities and a bulk temperature of 240 K.

The delay time of the second reflection allows estimating the thickness of the overlying material. The results indicate that basal reflector is approximately level with the surrounding plains. This excludes a significant downward deflection of the plains due to the load of the polar cap. A deflection of <500 m as indicated by the accuracy of the MARSIS measurements would put a lower limit of at least 150 km on the elastic thickness of the lithosphere (Johnson et al., 2000; Phillips et al., 2008).

10.11.3.7 Large Debris Flows

A large variety of gravity-driven flow or mass movement processes that involve water and sediments have been termed debris flows (see reviews by Coussot and Meunier (1996) and Iverson (1997)). Several landforms on Mars have been interpreted as the products of erosion and sedimentation by debris flows. While the small-scale gullies and the rock glacier-like features are discussed elsewhere in this chapter

(see Sections 10.11.3.8 and 10.11.3.5), here, we give a short overview of possible large-scale debris flow processes on Mars. Debris flows were invoked several times to explain the complex assemblages of collapsed surface regions and the huge outflow channels. For example, a debris flow mechanism was proposed to account for the formation of the chaotic terrains and the outflow channels debouching into Chryse Planitia (Nummedal and Prior, 1981). Huge smooth deposits on the floor of some outflow channels and extending into Chryse Planitia have been interpreted to be debris flows (Tanaka, 1997, 1999).

Large flow-like deposits northwest of the Elysium volcanic rise might have been emplaced as lahars (Christiansen, 1989; Russell and Head, 2003). Lahars are debris flows originating at volcanoes, often triggered by the melting of ice (glaciers and ground ice) by volcanic heat (e.g., Vallance, 2000). Water was also proposed to be a constituent in Martian landslides, mainly in Valles Marineris (Lucchitta, 1979, 1987). It has to be noted, however, that the presence of water in Martian landslides is still under debate, and a dry origin for the landslides in the Valles Marineris also seems possible (McEwen et al., 1989), a view that was recently confirmed by lab-scale dry granular flow experiments (Lajeunesse et al., 2006) and comparative studies with landslides on Earth (Soukhovitskaya and Manga, 2006).

It should be mentioned that interpreting debris flow features and deriving implications on their water content and emplacement mechanisms are not without problems on Mars. The strong and possible long-term modification of landforms on Mars by cold-climate processes often hinders the identification of their original nature (Van Gasselt et al., 2007).

10.11.3.8 Gullies

The discovery of gullies on Mars was one of the most exciting findings of the MGS mission because most gullies are pristine and not superposed by impact craters, which is a sign of their formation in recent geologic timescales. Because of their small scale, they were first detected in high-resolution MOC images (Malin and Edgett, 2000). The morphology of these features indicates the involvement of a fluid medium, most probably water, in their formation. As a consequence, the gullies are important for the Martian history because they may attest to very recent hydrologic activity on Mars.

10.11.3.8.1 Morphology and Occurrence

Gullies on Mars resemble terrestrial features formed by a combination of processes including overland and debris flow, which vary in water content and sediment supply (Figure 12). The Martian features show three typical morphological attributes that can be divided into (1) head alcove (source region), (2) channel (transport zone), and (3) debris apron (deposition zone) (Malin and Edgett, 2000). Gully sizes range in lengths from a few hundred meters to some kilometers and widths from some meters to several tens of meters. Morphological characteristics of the gullies such as anastomosing, braided, and sinuous channels with multiple

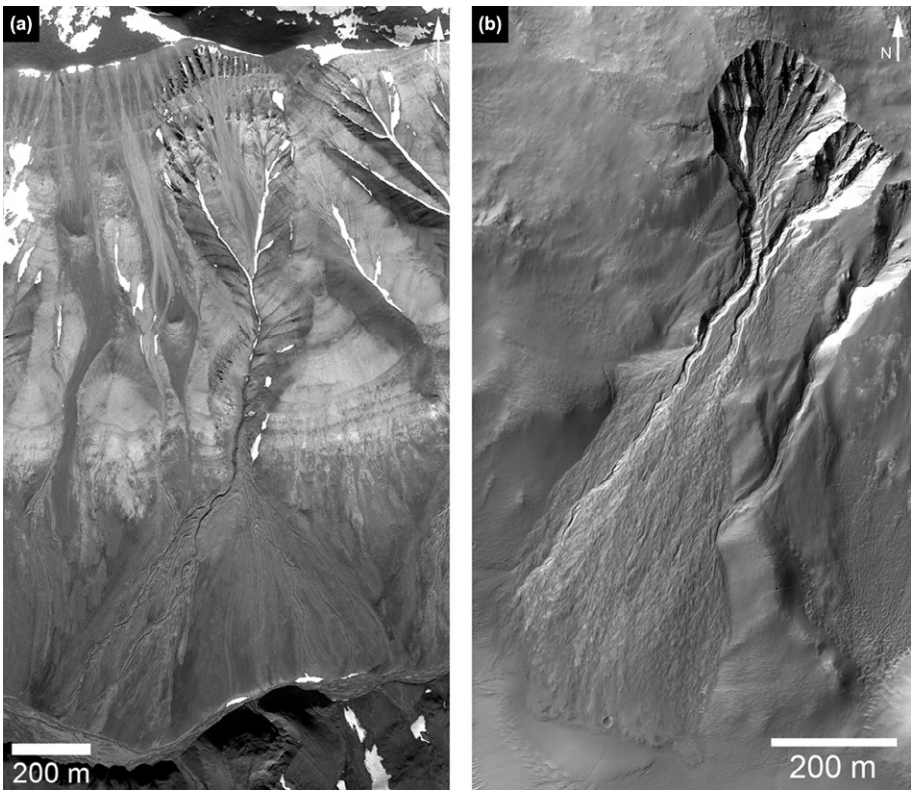


Figure 12 (a) Gully on Svalbard and (b) Mars. (a) High Resolution Stereo Camera (HRSC-AX) aerial image at 78.3° N and 15.8° E; 20 cm per pixel. (b) High Resolution Imaging Science Experiment (HiRISE) image PSP_006888_1410 at 38.5° S and 319.8° E; 25 cm per pixel. Image credit NASA/JPL/University of Arizona.

terraces, point bars, and cutbanks (e.g., Dickson and Head, 2009; Malin and Edgett, 2000; McEwen et al., 2007a,b) suggest that fluvial processes were involved in their formation. Modeling of Martian gully systems suggests that debris flows are the dominant process for gully formation on Mars (Conway et al., 2011; Mangold et al., 2010), and some Martian gullies show morphological evidence such as lateral ridges (levees) along channels, debris lobes or snouts, and debris plugs (Lanza et al., 2010; Levy et al., 2010; Mangold et al., 2010; Reiss et al., 2011). However, most gullies do not show clear evidence for debris flow processes possibly because debris flow characteristics were eroded or obscured by dust or fluvial processes were the dominant process in gully formation on Mars.

Gullies on Mars are tied to steep terrains. Most of them occur on crater slopes and, less frequently, on trough and valley slopes. Other locations include slopes of crater central peaks, isolated knobs, and dunes. The global distribution is limited to midlatitudes and high latitudes poleward of ~30° in both hemispheres but most frequently in the 30°–50° latitude bands (Balme et al., 2006; Bridges and Lackner, 2006; Dickson et al, 2007; Heldmann and Mellon, 2004; Heldmann et al., 2007; Kneissl et al., 2010; Malin and Edgett, 2001; Milliken et al., 2003). Gullies are much more common in the southern hemisphere. This asymmetry is probably due to the less cratered and smooth terrain in the northern hemisphere in contrast to the heavily cratered southern highlands with a high amount of steep slopes (Malin and Edgett, 2001). Global studies showed a preferred poleward-facing slope orientation of gullies (e.g., Malin and Edgett, 2000).

10.11.3.8.2 Formation

The formation mechanism for gullies is controversially debated since their discovery. The morphology of the gullies suggests that they are formed by a mass-wasting process behaving like a fluid. The fluid medium as well as the source of it is still discussed. Hoffman (2000b), Hoffman (2002), and Musselwhite et al. (2001) suggested that carbon dioxide from surface snow or the subsurface can carve the channels by gas-lubricated flows with entrainment of clastic debris. Additionally, Jöns (2002) suggested that the decomposition of CO_2 hydrates can deliver sufficient amounts of liquid water for the erosional processes. These theories seem unlikely because neither CO_2 nor CO_2 hydrates can accumulate in significant quantities in the subsurface and the morphology of possible gas-supported flows is not consistent with the observed gully morphology (Stewart and Nimmo, 2002). Granular flows are suggested by Treiman (2003) and Shinbrot et al. (2004), but dry mass movements cannot form the observed characteristic morphologies. However, there are a broad variety of different gully morphologies, and some Martian features may resemble dry mass movement features on the Moon (Bart, 2007). The most likely medium to explain the gully morphology is liquid water. Two main theories depending on the source of water exist. The water is suggested to be released from the subsurface or deposited from the atmosphere and melted by insolation. Malin and Edgett (2000) proposed a model in which a shallow subsurface reservoir of liquid water is protected against evaporation by an ice barrier. Sudden release of liquid water and subsequent debris flows occur when the icy plug is ruptured (Malin and Edgett, 2000). Gilmore and Phillips (2002) favored melting of near-surface ice that percolates through a permeable surface layer and subsequent seepage at the base of an impermeable layer in the subsurface. Others suggest that geothermal heating melts subsurface ice (Gaidos, 2001; Hartmann, 2001; Mellon and Phillips, 2001). Atmospheric-dependent models suggest that the water ice is deposited out of the atmosphere and melted by insolation during periods of high-obliquity phases of Mars. Costard et al. (2002) suggested accumulation of ice within the regolith and melting of this near-surface ice by insolation during periods of high obliquity. In this model, the near-surface ice is protected against evaporation by seasonally deposited CO_2 frost. Lee et al. (2001) and Christensen (2003) proposed a model in which water ice is deposited as snow during high-obliquity periods and melted at low obliquity. Here, melting occurs beneath the snow cover sheltering the liquid water against evaporation (Christensen, 2003). Head et al. (2008) proposed a top-down melting of residual ice and snow after the last glaciation.

The morphological characteristics of gullies, their limited global distribution in midlatitudes and high latitudes, and their preferred orientation on poleward-facing slopes are in agreement with water being involved in gully formation. The model of atmospherically derived ice or snow deposits is consistent with other morphological observations of a deposited dust–ice mantle and viscous flow features in the midlatitudes (Milliken et al., 2003; Mustard et al., 2001). Several gullies also emanate from the dust–ice mantle (e.g., Arfstrom and Hartmann, 2005; Aston et al., 2011; Berman et al., 2005; Bleamaster and Crown, 2005; Christensen, 2003; Raack et al., 2012), suggesting that melting from the dust–ice mantle formed these gullies. The formation of an ice-rich mantle, viscous flow features, and gullies are possibly all related to recent ice ages within the last few million years on Mars. According to this theory, water ice was removed from the poles and transported and deposited as a dust–ice mantle to midlatitudes during higher obliquities, which desiccated during lower obliquities (Head et al., 2003a). Modeling results indicate that centimeters of snow can be seasonally deposited and melted in midlatitude regions on Mars under high-obliquity (>35°) conditions in the past (e.g., Mischna et al., 2003; Williams et al., 2009).

10.11.3.8.3 Age

The gullies are suggested to be formed geologically recently. Estimates that the gullies could be < 1 Ma years old are based on their pristine appearance, their stratigraphic relationships to young surface features, and the lack of superimposed impact craters (Malin and Edgett, 2000). Other age estimations assume that the gullies were formed under different climatic conditions than today and correlate their activity to obliquity variations in the last few million years (e.g., Christensen, 2003; Costard et al., 2002; Head et al., 2003a; Lee et al., 2001). Crater counts on dunes superposed by gully deposits in Nirgal Vallis indicate that the gullies were formed within the last 3 Ma, possibly within the last 300 ka (Reiss et al., 2004). Schon et al. (2009) dated a crater with 0.6–2.4 Ma (best fit 1.25 Ma) that produced secondary craters superposing gully sedimentary fans. Although most of the gullies appear morphologically

pristine, there is observational evidence of older gullies that are degraded (filled channels, eroded aprons, and cut by fissures) or cratered (e.g., McEwen et al., 2007a,b; Morgan et al., 2010a,b; Reiss et al., 2009). These older examples indicate that not all gullies are young and their formation occurred at least episodically over a longer timescale (Morgan et al., 2010a,b, 2011; Reiss et al., 2009).

10.11.3.8.4 Dune gullies

Gullies have also been found in several places on slopes of midlatitude dark dunes in the southern hemisphere that can be classified into two main morphologies (Reiss et al., 2007):

(1) Long sinuous channels, with a constant width from the source region to their termini; lateral ridges (levees) along the channels; and debris snouts at their termini (lack of sedimentary fan), also called 'linear dune gullies' (Védie et al., 2008) (**Figure 13(a)**).
(2) Dune gullies with an alcove, channel, and sedimentary fan resembling the typical morphology of midlatitude slope gullies elsewhere on Mars, also called 'ACA (alcove–channel–apron) dune gullies' (Dundas et al., 2012) (**Figure 13(b)**). Linear dune gullies occur on several slopes of midlatitude dark dunes (e.g., Costard et al., 2002; Mangold et al., 2003; Reiss and Jaumann, 2003). Based on their morphological characteristics, flow rheology, and laboratory simulations, they were interpreted to be formed by debris flow (mudflows and slurry flows) with the involvement of about >10% by volume of liquid water (Jouannic et al., 2012; Mangold et al., 2003; Miyamoto et al., 2004; Reiss and Jaumann, 2003; Reiss et al., 2010; Védie et al., 2008). An alternative formation mechanism carving the channels by mass movements of CO_2 blocks was proposed by Dundas et al. (2012) and Diniega et al. (2013). ACA dune gullies were interpreted to be formed by H_2O- or CO_2-related processes (Reiss et al., 2007) and recently more to be caused by CO_2 processes (Diniega et al., 2010; Dundas et al., 2012), see **Section 10.11.3.8.5**.

10.11.3.8.5 Present-day activity

Bright sedimentary deposits (**Figure 13(c)**) were observed to form under present-day conditions at gully sites and interpreted to be caused by flows involving liquid water (Malin et al., 2006). Based on analogue studies, Hugenholtz (2008) proposed that the bright deposits on Mars might be formed by vaporization of CO_2–ice small amounts of H_2O–ice lubricating fine-grained material. The following studies did not rule out the presence of liquid water, but indicate that dry flows can form all the morphological characteristics of the bright deposits (Kolb et al., 2010; Pelletier et al., 2008). Multitemporal image analysis showed that the bright deposits do not change their tone or shape (Kolb et al., 2010; McEwen et al., 2007a,b), which would be expected when the bright deposits would contain water or ice (Williams et al., 2007). In addition, ice or hydrated minerals could not be detected with hyperspectral data (McEwen et al., 2007a,b). However, spectral analysis of bright deposits left by former mudflows in the Atacama Desert showed no hydrated minerals suggesting that liquid water cannot be ruled out as a possible formation hypothesis for the bright gully deposits (Heldmann et al., 2010).

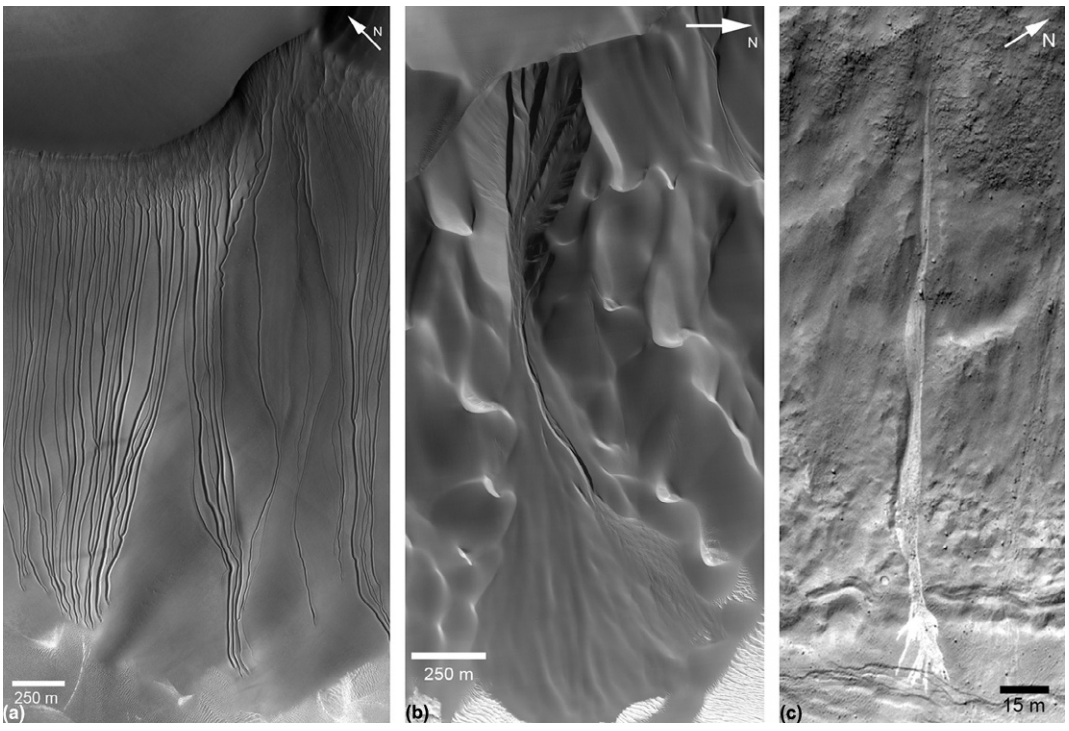


Figure 13 (a) Linear dune gullies in Russell Crater at 54.3° S and 13° E (HiRISE image PSP_007018_1255). (b) Alcove–channel–apron (ACA) dune gullies in Matara Crater at 49.5° S and 34.9° E (HiRISE image PSP_006648_1300). (c) Bright sedimentary flows in Terra Sirenum at 36.3° S and 198.3° E (HiRISE image PSP_003596_1435). Image credit NASA/JPL/University of Arizona.

Recently, new multitemporal observations revealed modifications of several gully channels and aprons including widening of channel segments, possible channel incision, and new dark deposits on the aprons (Dundas et al., 2010, 2012). The changes occur in the winter seasons and are interpreted to be caused by CO_2 frost and some by dry granular flows (Dundas et al., 2010, 2012). Present-day activity of dune gullies was also observed. ACA dune gullies show morphological changes of alcove and channel widening and new depositions on aprons (Diniega et al., 2010; Dundas et al., 2012). Based on the seasonal behavior of the ACA dune gully activities, their formation is suggested to be triggered by CO_2-driven processes (Diniega et al., 2010; Dundas et al., 2012). Reiss et al. (2010) observed present-day activity of linear dune gullies where a small ~2 m wide channel grew by incision into the dune several tens of meters in length in two subsequent years. The activity was interpreted to be caused by melting of small amounts of H_2O–ice triggering slurry flows of sand mixed with liquid water based on seasonal spectral data and modeled surface temperatures (Reiss et al., 2010).

10.11.3.9 Recurring Slope Lineae

Recurring slope lineae (RSL) are narrow (up to 5 m wide), relatively dark-toned features on steep (>25°) equator-facing slopes in the southern hemisphere between 48° and 32° S latitude (**Figure 14**) (McEwen et al., 2011). They appear and lengthen in late spring and summer and slowly fade and disappear during autumn and winter (McEwen et al., 2011). Their seasonal behavior, latitudinal occurrence, preferred slope orientation, and small size suggest that these features could result from brine flows (McEwen et al., 2011). Terrestrial analogue studies in Antarctica combined with a groundwater flow model used for the Martian and terrestrial RSL analogues support the hypothesis that RSL form through the downslope flow of liquids (Levy, 2012). Chevrier and Rivera-Valentin (2012) used a combination of thermodynamic and kinetic numerical models to investigate if the RSL can be formed by melting of brines. Their model results reproduced well the seasonality and preferred slope orientations (Chevrier and Rivera-Valentin, 2012).

10.11.4 Water in Mantle and Crust

Looking at the terrestrial planets today, Mars is the only planet on which surface water has been detected directly in the form of ice at the polar caps (Bibring et al., 2004a; Hansen et al., 2005). There is also evidence for large amounts of ground ice in the circumpolar areas (Feldman et al., 2004). Recent radar observations by the MARSIS instrument on Mars Express (Picardi et al., 2005) and theoretical modeling indicate the existence of ground-ice deposits extending to midlatitudes and low latitudes (Helbert et al., 2005). For Mercury (Butler et al., 1993 and Harmon and Slade, 1992) as well as for the Moon (Feldman et al., 1998; Nozette et al., 1996), there are indications for the existence of ice deposits in the regolith in permanently shadowed areas in the polar regions (see

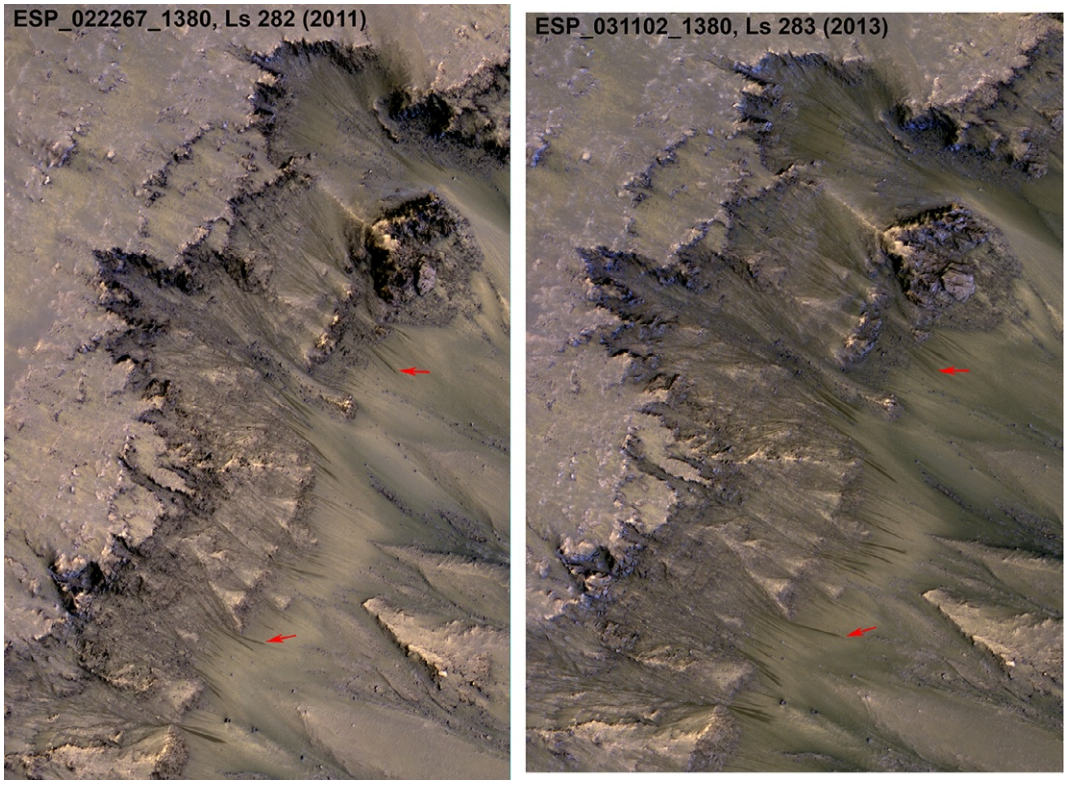


Figure 14 Seasonal changes of recurring slope lineae (RSL) in Palikir Crater at 41.6° S and 202.3° E during one martin year. Image credit: NASA/JPL/University of Arizona.

Section 10.11.4.1.1). These deposits are still subject to debate and await confirmation by future missions. The present-day conditions clearly exclude the existence of liquid water or ice on the surface of Venus.

10.11.4.1 Regolith

10.11.4.1.1 Moon and Mercury

The rotation axis of the Moon has an inclination of only 1.6° resulting in areas near the poles, primarily the floors and lower interior walls of impact craters, that are permanently shadowed from the Sun. It was suggested very early (Arnold, 1979; Watson et al., 1961) that temperatures within these permanently shadowed regions near the lunar poles were sufficiently low that water would be stable to sublimation for the lifetime of the Moon. Meteoroids and comets were proposed as source for the water. A number of follow-up studies refined this idea (Butler, 1997; Butler et al., 1993; Vasavada et al., 1999). The first tentative observation indicating possible ice at the lunar south pole was reported by Clementine from bistatic radar observations of very low spatial resolution (Nozette et al., 1996). The 13 cm wavelength telemetry system of the Clementine orbiter was used as a transmitter and one of the 70 m NASA/Deep Space Network antennas to receive the echo. The signal received from the pass over the south pole showed a lower depolarization as the other passes indicate the presence of a low-loss volume scatterer, possibly water ice. A follow-up analysis by Nozette et al. (2001) suggested a body of ~ 10 km^2 of ice in the Earth-facing lower part of Shackleton crater. However, analysis of Arecibo radar data (Stacey et al., 1997) and a reanalysis of the Clementine data (Simpson and Tyler, 1999) have raised questions regarding this conclusion. An independent indication for possible ice deposits at the poles of the Moon was provided by an analysis of Lunar Prospector (LP) neutron data. These data showed enhanced abundances of hydrogen, existing at both poles of the Moon (Feldman et al., 1998, 2000). The spatial resolution of the LP measurements was not sufficiently good to prove a one-to-one correspondence of the signal to the permanently shaded craters near both poles (Bussey et al., 1999; Margot et al., 1999). However, it was possible to show that an association was plausible. Using models of neutron emission from assumed water-ice deposits filling the largest three craters in the neighborhood of the south pole, a water-ice abundance of $1.5 \pm 0.8\%$ by mass was derived. The result is however not a proof for the existence of water ice, because the measurement of hydrogen is independent of its chemical association. Recent high-spatial-resolution Earth-based radar observations using the 13 cm radar imaging with the Arecibo radar system showed no evidence for characteristic changes in the depolarization in Shackleton crater or anywhere else near the south pole of the Moon (Campbell et al., 2006). These results in combination with the LP results implied a low likelihood for a thick ice deposit within the crater but rather favor ice being dispersed within the regolith.

The distribution of hydrogen across the region is consistent with buried water ice from cometary impacts, hydrogen implantation from the solar wind, and/or other as yet unknown sources.

IR spectroscopic measurements of the lunar surface from three different spacecraft published in 2009 provide unambiguous evidence for the presence of hydroxyl (OH) or water on the surface of the Moon. The Moon Mineralogy Mapper (M3) on Chandrayaan-1 detected absorption features near 2.8–3.0 μm on the surface of the Moon (Pieters et al., 2009). Based on the general lack of correlation of this feature in sunlit M3 data with neutron spectrometer H abundance data, Pieters et al. (2009) suggested that the formation and retention of OH and H_2O is an ongoing surficial process. OH/H_2O production processes may feed polar cold traps. The Deep Impact spacecraft found the entire surface to be hydrated during some portions of the day (Sunshine et al., 2009). OH and H_2O absorptions in the near-IR were strongest near the north pole and are consistent with <0.5 wt% H_2O. Hydration varied with temperature, rather than cumulative solar radiation, but no inherent absorptivity differences with composition were observed. However, comparisons between data collected one week (a quarter lunar day) apart show a dynamic process with diurnal changes in hydration that were greater for mare basalts ($\sim$70%) than for highlands ($\sim$50%). This hydration loss and return to steady state occurred entirely between local morning and evening, requiring a ready daytime source of water group ions, which is consistent with a solar wind origin. Finally, looking at data from the Visual and IR Mapping Spectrometer on Cassini during its flyby of the Moon in 1999, Clark (2009) observed a broad absorption at 3 μm due to adsorbed water and near 2.8 μm attributed to hydroxyl in the sunlit surface on the Moon. The amounts of water indicated in the spectra depend on the type of mixing and the grain sizes in the rocks and soils but could be 10–1000 ppm and locally higher. Water in the polar regions may be water that has migrated to the colder environments there. Trace hydroxyl is observed in the anorthositic highlands at lower latitudes.

The Lunar Crater Observation and Sensing Satellite (LCROSS) mission was designed to detect hydrogen-bearing volatiles directly (Colaprete et al., 2010). On 9 October 2009, a spent Centaur rocket struck the persistently shadowed region within the lunar south pole crater Cabeus, ejecting debris, dust, and vapor. A second 'shepherding' spacecraft, which carried nine instruments, including cameras, spectrometers, and a radiometer, observed this material. Near-IR absorbance attributed to water vapor and ice and UV emissions attributable to hydroxyl radicals support the presence of water in the debris. The maximum total water vapor and water ice within the instrument field of view was 155 ± 12 kg. Given the estimated total excavated mass of regolith that reached sunlight, and hence was observable, the concentration of water ice in the regolith at the LCROSS impact site is estimated to be $5.6 \pm 2.9\%$ by mass. In addition to water, spectral bands of a number of other volatile compounds were observed, including light hydrocarbons, sulfur-bearing species, and carbon dioxide.

Neutron flux measurements of the Moon's south polar region from the Lunar Exploration Neutron Detector (LEND) (Mitrofanov et al., 2010) on the Lunar Reconnaissance Orbiter spacecraft were used to select the optimal impact site for LCROSS. LEND data show several regions where the epithermal neutron flux from the surface is suppressed, which is indicative of enhanced hydrogen content. These regions are not spatially coincident with permanently shadowed regions of the Moon. The LCROSS impact site inside the Cabeus crater demonstrates the highest hydrogen concentration in the lunar south polar region, corresponding to an estimated content of

0.5–4.0% water ice by weight, depending on the thickness of any overlying dry regolith layer.

The same arguments made for the Moon are valid for Mercury as well. Thermal modeling (Ingersoll et al., 1992; Paige et al., 1992, 2013) showed that ice can remain stable over geologic timescales in permanently shadowed areas near the poles of Mercury. This was confirmed by Earth-based radar observation discovering probable ice deposits at the poles of Mercury (Harmon and Slade, 1992; Slade et al., 1992). The radar echoes from Mercury's poles were interpreted as reflections from water ice on the basis of their similarity to the unusual properties of radar echoes from the icy Galilean satellites of Jupiter (Campbell et al., 1978). Higher-spatial-resolution radar images (see **Figure 15**), obtained by Harmon et al. (2001), confirmed all the original polar features and revealed many additional features, including several at latitudes as low as 72–75° N and several from craters <10 km in diameter. All of the new features located on the Mariner-imaged side of the planet could be matched with known craters or other shaded areas. Moses et al. (1999) showed by extrapolating the current terrestrial influx of interplanetary dust particles to that at Mercury that continual micrometeorite bombardment of Mercury over the last 3.5 billion years could have resulted in the delivery of $(3–60) \times 10^{13}$ kg of water ice to the permanently shaded regions at Mercury's poles. This would result in an average ice thickness of 0.8–20 m. Erosional processes like solar wind sputtering and micrometeorite bombardment would remove about half of this ice in the same time span. The resulting amount of ice is in general agreement with the proposed amount of $\sim$2–20 m based on radar imaging if the additional influx by comets and asteroids is taken into account. Butler et al. (1993) discussed the notion that water originating away from the poles would behave as individual, randomly moving molecules, some of which could migrate to the poles and become cold-trapped there.

Observations by the NASA MESSENGER spacecraft confirmed this notion (Lawrence et al., 2013). The neutron spectrometer on the MESSENGER spacecraft shows decreases in the flux of epithermal and fast neutrons from Mercury's north polar region that are consistent with the presence of water ice in permanently shadowed regions. The neutron data indicate that Mercury's radar-bright polar deposits contain, on average, a hydrogen-rich layer more than tens of centimeters thick beneath a surficial layer 10–20 cm thick that is less rich in hydrogen. The buried layer must be nearly pure water ice. The upper layer contains <25 wt% water-equivalent hydrogen. The total mass of water at Mercury's poles is inferred to be 2×10^{16}–10^{18} g and is consistent with delivery by comets or volatile-rich asteroids. This discovery was independently confirmed by measurements from the laser altimeter on the MESSENGER spacecraft (Neumann et al., 2013). Measurements of surface reflectance of permanently shadowed areas near Mercury's north pole reveal regions of anomalously dark and bright deposits at 1064 nm wavelength. These reflectance anomalies are concentrated on poleward-facing slopes and are spatially collocated with areas of high radar backscatter postulated to be the result of near-surface water ice. Correlation of observed reflectance with modeled temperatures indicates that the optically bright regions are consistent with surface water ice, whereas dark regions are consistent with a surface layer of complex organic material that likely overlies buried ice and provides thermal insulation.

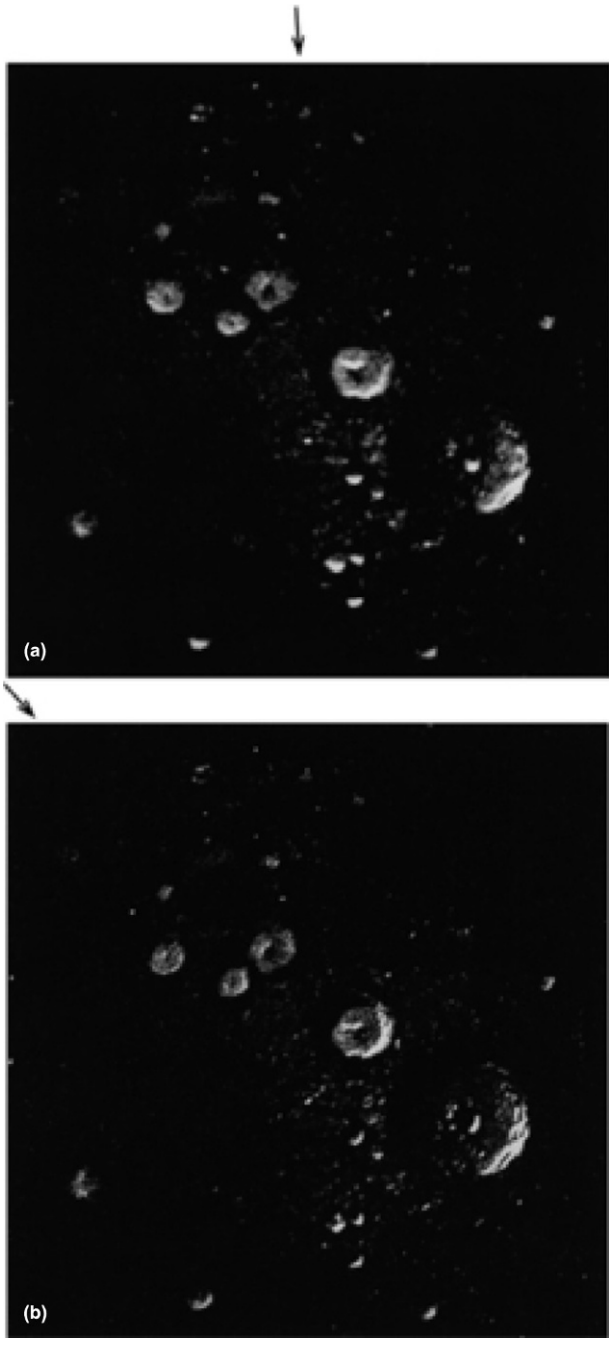


Figure 15 Radar images of the deposit close to the pole of Mercury. Lighter shades correspond to higher echo strength. The radar illumination directions are indicated (arrows). Surface resolution is 3 km in (a) and 1.5 km in (b). Reproduced from Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's North Pole. *Icarus* 149: 1–15.

10.11.4.1.2 Mars

The stability of water ice on the surface of Mars is governed by the balance between sublimation and deposition and mainly a function of the atmospheric temperature and water vapor

abundance. For subsurface water-ice deposits in the regolith, the situation is more complex, because the exchange rate with the atmosphere is controlled by the physical properties of the regolith. In principle, two end-member situations can be identified. One extreme scenario is that the subsurface ice is disconnected from the atmosphere by an impermeable layer of material. The other extreme would be a free exchange of water vapor between the subsurface ice deposit and the atmosphere. While the regolith would protect the subsurface ice deposit from temperature variation, ground ice and atmospheric vapor would eventually reach equilibrium (Flasar and Goody, 1976; Mellon and Jakosky; 1993; Schorghofer and Aharonson, 2005). Between these two extremes, there are a variety of other possibilities. Helbert et al. (2005) proposed, for example, that the retreating ground ice forms a sublimation till, which limits the vapor transport to the atmosphere, thereby stabilizing the ice.

The wide range of assumptions on the interaction between subsurface ice and the atmosphere lead to a disagreement about the stability of proposed ground-ice deposits especially at midlatitudes and low latitudes. The calculations by Farmer and Doms (1979), Mellon and Jakosky (1993), and Mellon and Jakosky (1995) show that water ice must be unstable at shallow depths for latitudes $<65°$. Schorghofer and Aharonson (2005) took into account seasonal variations of air moisture in their model and found that the ice can be stable under some conditions, in particular, at low thermal inertia. In the model by Helbert et al. (2005), subsurface ice is metastable at these depths: it disappears, but slowly over geologically significant timescales. On Mars, the variation of climate over the history of the planet plays an important role in the distribution and amount of ice in the regolith.

Impact cratering into ice-rich (and possibly fluid) target material is suggested to form a specific layered ejecta morphology on Mars termed rampart crater, which is not observed on other terrestrial planets. Models suggest that the target material must contain at least 15–20% ice for the formation of a layered ejecta blanket (Stewart et al., 2001; Woronow, 1981). The ejecta are characterized by 'lobate'- or 'fluidized'-like structures, which terminate in distal ramparts (**Figure 16**). A summary of different layered ejecta morphologies is given in Barlow et al. (2000). Although most researchers suggest impacts into volatile-rich target material for their formation (e.g., Carr et al., 1977; Head and Roth, 1976; Mouginis-Mark, 1979), atmospheric processes might also cause or contribute to it (e.g., Barnouin-Jha and Schultz, 1998; Schultz, 1992; Schultz and Gault, 1979). For a summary of models, morphological constraints, laboratory experiments, and discussion about the two formation theories, see the reviews by Barlow (2005) and Kuzmin (2005).

Rampart craters are found in all geologic units and at all altitudes and latitudes (Allen, 1979; Barlow and Bradley, 1990; Costard, 1989; Kuzmin et al., 1988; Mouginis-Mark, 1979). In a given area, a certain minimum diameter exists for rampart craters, called the onset diameter (Boyce and Witbeck, 1980; Kuzmin, 1980). Geographic mapping revealed that the onset diameter decreases toward higher latitudes (Costard, 1989; Kuzmin et al., 1988; Mouginis-Mark, 1979). Rampart crater diameters in equatorial regions are typically 4–7 km versus 1–2 km at high latitudes (50° latitude) (Squyres et al., 1992).

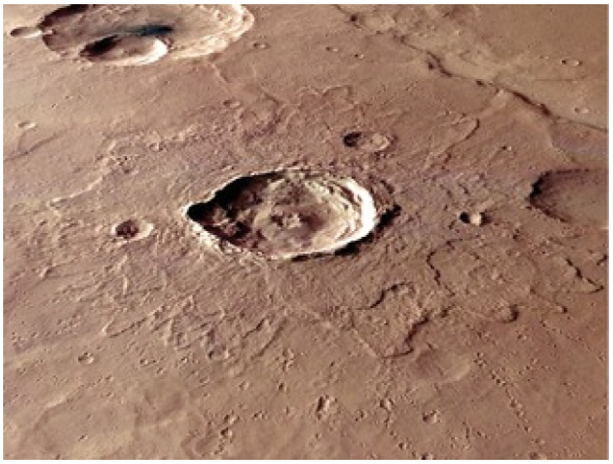


Figure 16 Perspective color view of a rampart crater (diameter = 9.5 km) in southern Arabia Terra, Mars (HRSC orbit 2956_0000 at 16.5° E and 15.2° N). copyright: ESA/DLR/FU Berlin.

Based on the physics of the impact process and using a relationship between the onset diameter and the crater's excavation depth, Kuzmin et al. (1989) derived global depths of the boundary between a desiccated regolith layer and an ice-bearing regolith layer. The results indicate a possible boundary in ~300 to ~400 m near the equator and ~100 m at 50° latitudes (Kuzmin, 2005). Barlow et al. (2001) showed that the onset diameter varies regionally near the equator, indicating shallower ground-ice depths in these regions. Reiss et al. (2005) found regionally equatorial onset diameters of rampart craters as small as 1 km, equal to those at high latitudes. However, the layered ejecta blankets are highly degraded, indicating a formation in the early history of Mars and possible shallow ground-ice depths at these times (Reiss et al., 2005). Rampart craters may have formed over a significant time interval and the onset diameter records the ground-ice depth at the time of the impact (Squyres et al., 1992). Possible changes with time might be detectable. Barlow (2004) analyzed characteristics of layered ejecta in equatorial regions and found no major variations in volatile concentrations at depths >1 km. Based on absolute age determinations by crater counts on the layered ejecta of rampart craters in equatorial regions and in comparison with the onset diameter, Reiss et al. (2006) suggested a decrease of the ground-ice table at depths <1 km in these regions.

10.11.4.2 Mineralogy

Minerals are the basic rock-forming constituents of planetary surfaces. As such, their investigation can provide valuable information about the aqueous history of a planet. The most interesting minerals in this respect are hydrated minerals, that is, minerals that include water (H_2O) or hydroxyl (OH) in their mineral structures. Examples of hydrated minerals are iron hydroxides, carbonates, sulfates, phyllosilicates, and zeolites. The only terrestrial planets that have been found to have hydrated minerals at their surfaces are the Earth and Mars. Some asteroids also possess OH- and/or H_2O-bearing minerals (Rivkin et al., 2002), for example, Vesta (Hasegawa et al., 2003),

but these are not further discussed here. On the other hand, there is no evidence for water-containing minerals on the Moon (Lucey et al., 2006), or on Venus, although on Venus too little of the mineralogy is known to make a firm statement. It is not in the scope of this chapter to discuss hydrated minerals on Earth. Summaries of Martian surface mineralogy, including potential hydrated minerals, are given by Soderblom (1992), Roush et al. (1993), Bell (1996), and Bibring and Erard (2001). Bish et al. (2003) discussed the stability of hydrous minerals on Mars, Bishop (2005) gave a recent but still pre-MER and pre-MEX overview of the properties of hydrated minerals on Mars, and a review of the several forms of hydration that can exist on the Martian surface is presented by Jouglet et al. (2006). Over the last decade, the results obtained by spectrometers in orbit (OMEGA and CRISM) and in situ on board the MER and MSL rovers have dramatically changed the picture for Mars, and some of the most relevant results are presented in the succeeding text.

10.11.4.2.1 Methods

Typically, the composition of Martian surface materials has been analyzed on the basis of Earth-based spectroscopic (e.g., McCord and Adams, 1969), in situ chemical (e.g., Clark et al., 1977; Rieder et al., 1997; Wänke et al., 2001), and in situ magnetic (e.g., Hargraves et al., 1977) data (see Bell, 2008). Reflectance spectroscopy is a particularly powerful tool to infer the mineralogy of planetary surfaces (e.g., Burns, 1970), but only few instruments were operated in orbit around Mars. The near-IR spectrometer (IRS) aboard Mariner 6 and Mariner 7 (Calvin, 1997) and imaging spectrometer for Mars (ISM) aboard Phobos-2 (Bibring et al., 1989 and Murchie et al., 1993) detected a broad and deep 3 μm absorption band in each spectrum. The strength and shape of the 3 μm band are influenced by all forms of hydration. Unfortunately, the ISM acquired only a very limited number of observations before the Phobos 2 mission failed at an early stage. The most recent near-IR imaging spectrometers in orbit around Mars are OMEGA on board Mars Express (Bibring et al., 2004b) and the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on board the MRO (Murchie et al., 2009). The Thermal Emission Spectrometer (TES) on board the MGS Mission (Christensen et al., 1992) and the THEMIS instrument on board the Mars Odyssey mission (Christensen et al., 2004a) both acquired spectra at longer wavelengths in the thermal emission range.

The MERs, Spirit and Opportunity, have several instruments that can provide mineralogical information (Squyres et al., 2003). The panoramic camera (Bell et al., 2003) and the miniature thermal emission spectrometer (Christensen et al., 2003) provide multispectral data and can still be considered to be 'remote-sensing' experiments, despite their very close range to the observed targets. The Mössbauer spectrometer (Klingelhöfer et al., 2003) and the alpha particle x-ray spectrometer (APXS) (Rieder et al., 2003), which provides geochemical rather than mineralogical information, take their measurements in direct contact with the target materials. A variety of hydrated minerals have been identified on Mars (see succeeding text; e.g., Bibring et al., 2006; Carter et al., 2013; Ehlmann et al., 2013).

10.11.4.2.2 Basalt and dust

Overall, the surface of Mars consists of a mixture of relatively pristine mafic materials (i.e., basalt consisting of pyroxenes and olivine) and highly oxidized secondary materials (Bell, 1996). Mössbauer and APXS data collected by the MERs show that the ubiquitous, oxidized bright red dust, which is responsible for the red color of Mars, consists of ferric oxides and includes nanocrystalline phases (e.g., Goetz et al., 2005). It is an alteration product, probably of a primary material with basaltic chemistry, but the presence of olivine in the dust, which is quickly weathered in the presence of water, suggests that water did not play a significant role in its production (Goetz et al., 2005). This is also confirmed by orbital observations with the OMEGA instrument. It found evidence for widely distributed anhydrous ferric oxides in high-albedo regions (Bibring et al., 2006), which might consist of nanocrystalline red hematite (α-FeO) and maghemite (γ-FeO). Therefore, both rover and orbiter data point to the formation of the Martian dust that occurred under a dry climate. Nevertheless, recent observations by the Curiosity rover show that dust contains a significant amount of H (Lasue et al., 2013a,b; Meslin et al., 2013), which may account for much of the hydrogen detected by the GRS from orbit (Feldman et al., 2004).

10.11.4.2.3 Hematite

Spectra taken with TES show evidence for gray, coarse-grained (>10 μm) crystalline hematite in three limited regions, Aram Chaos, Meridiani Planum, and, to a lesser degree, Valles Marineris (**Figure 17**) (Christensen et al., 2000, 2001b). Hematite is not a hydrated mineral, but it is thought to have formed via aqueous processes. Although several mechanisms exist to form this variety of hematite, Christensen et al. (2001b) favored chemical precipitation from Fe-rich aqueous fluids under either ambient or hydrothermal conditions (see also Catling and Moore, 2003). Such a process would have required substantial amounts of liquid water. From a morphological perspective, this appears plausible, since all three hematite deposits detected with TES are located in sedimentary settings.

Figure 17 Hematite in Meridiani Planum as mapped with the TES (e.g., Christensen et al., 2000). The hematite-rich areas (marked in red) have probably experienced long-lasting aqueous processes and were selected as landing site for one of the MERs. Image credit: NASA Planetary Photojournal, PIA02056.

Actually, the hematite deposit in Meridiani Planum was considered to be such strong evidence for past surface or near-surface water that one of the MERs, Opportunity, was sent there to explore its details (Golombek et al., 2003).

Indeed, the findings of the rover confirmed the presence of hematite in Meridiani Planum. Over the entire region, small millimeter-sized spherules (called 'blueberries' by the rover team) are embedded in the sandstones, which form the layered outcrops in Meridiani Planum. These spherules were found to consist of hematite, based on the observations of the Mössbauer and Mini-TES instruments (e.g., Christensen et al., 2004c; Klingelhöfer et al., 2004). Their accumulation in topographic depressions as a result of weathering is responsible for the detection of hematite from orbit.

10.11.4.2.4 Phyllosilicates

Phyllosilicates form a large group of minerals that includes, among others, micas, smectites, and kaolinites. Their basic characteristic feature is the layered structure, and their formation can take place at the surface (e.g., in soils or sediment) or subsurface (diagenesis and hydrothermal alteration) (e.g., Meunier, 2005). Some of them, for example, smectites, form as alteration products of volcanic material. Therefore, their presence on Mars has been considered possible and even likely on theoretical grounds (e.g., Bishop, 2005). Several studies suggested the existence of phyllosilicates on Mars based on in situ Viking lander analyses (Toulmin et al., 1977), the identification of smectites in laboratory analyses of shergottite, nakhlite, and chassigny (SNC) meteorites (e.g., Bridges et al., 2001), and spectroscopic measurements (e.g., Bishop et al., 1995; Clark et al., 1990; Wyatt and McSween, 2002). The general opinion, however, confirmed by the detection of sulfates in orbiter and rover measurements, was that the Martian mineralogy was dominated by acid sulfate processes, which prohibit the formation of clay minerals (see the comment by Newsom (2005)). Nevertheless, phyllosilicates were identified on Mars in near-IR spectra taken with the OMEGA spectrometer (Bibring et al., 2005). Clay minerals such as smectites were first identified from orbit in two major outcrops in the area of Mawrth Vallis and Nili Fossae and in several smaller-sized areas (Poulet et al., 2005). The phyllosilicate-rich materials were located in ancient (Noachian) regions, but their surfaces do not show large abundances of impact craters and, therefore, must have been exposed relatively recently. Later orbital observations added thousands of phyllosilicate-bearing surface materials (e.g., Carter et al., 2013; Ehlmann et al., 2013; Mustard et al., 2008), and even though most of them were found in ancient surfaces, some phyllosilicate formation appears to have continued into the Hesperian (Thollot et al., 2012).

Phyllosilicates have also been found in Gusev crater at the base of the Columbia Hills. Measurements of the Mössbauer and APXS spectrometers indicate the presence of members of the kaolinite, serpentine, chlorite, and septechlorite groups (Wang et al., 2006). The detection of Al-rich phyllosilicates suggests a mildly acidic past environment and an open hydrologic system with good drainage conditions (Wang et al., 2006). When both Fe/Mg- and Al-rich phyllosilicates are found together, the common stratigraphic relationship is that Al phyllosilicates overlie Fe/Mg phyllosilicates. This has been interpreted as a surface weathering sequence (e.g., Le Deit et al., 2012). On the other hand, it appears likely that no single process can explain all phyllosilicate outcrops on Mars, and multiple phyllosilicate-forming processes operated in diverse chemical environments, varying over space and time (Ehlmann et al., 2013).

10.11.4.2.5 Sulfates

Sulfates are salts of sulfuric acid and often contain water molecules. They are particularly useful to constrain the climatic evolution, since some sulfates can be transformed to other sulfates that have a different water content. While this reaction is sometimes reversible, it is irreversible for others. As an example, kieserite ($MgSO_4 \cdot H_2O$) can easily be hydrated and converts to hexahydrite ($MgSO_4 \cdot 6H_2O$) and epsomite ($MgSO_4 \cdot 7H_2O$). When these minerals are desiccated, they do not convert back to kieserite, but to an amorphous phase (Vaniman et al., 2004). The presence of kieserite at certain locations on Mars (see succeeding text) indicates, therefore, that the climate probably never experienced obliquity-driven periods of hydration and desiccation in these regions (Vaniman et al., 2004).

The presence of sulfates on Mars has been discussed for a long time. Theoretical considerations, partly based on terrestrial analogue materials, suggested that Mars might host sulfates, and, for example, Burns (1987) predicted that jarosite might be the result of the interaction of groundwater with iron sulfides. Viking lander measurements indicated a high sulfur content in the Martian soil, and Toulmin et al. (1977) proposed that sulfates might be the corresponding minerals. The SNC meteorites, which are thought to be samples of the Martian crust (e.g., Bogard and Johnson, 1983), contain sulfates (e.g., McSween, 1994), and spectroscopic observations also hinted toward sulfates (e.g., Blaney and McCord, 1995; Pollack et al., 1990).

Major sulfate deposits have been found with OMEGA in the Valles Marineris and Terra Meridiani (Gendrin et al., 2005) and the north polar region (Langevin et al., 2005a). In Valles Marineris, they are associated with layered deposits (**Figure 18**), and in the north polar region, they occur within a huge field of dark dunes. The MER confirmed the presence of sulfates and identified the Fe sulfate jarosite, predicted to exist on Mars by Burns (1987), in Meridiani Planum (Klingelhöfer et al., 2004). Most recently, the rover Curiosity identified Ca-sulfates in Gale crater, pointing to a more clement environment than the Fe and Mg sulfates found elsewhere (Nachon et al., 2013).

10.11.4.2.6 Carbonates

A warmer and denser greenhouse atmosphere on early Mars, as suggested by the geomorphic record of fluvial and lacustrine processes (see **Section 10.11.3.1**), would raise the question where the greenhouse gases have disappeared since then. A plausible way to remove CO_2 from the atmosphere would be the formation of carbonates, and the search for carbonates on Mars has been the subject of many investigations (Bibring et al., 2005; Christensen et al., 2001a). In fact, carbonates were identified in SNC meteorites (e.g., Bridges et al., 2001) and in surface materials both from orbit and in situ (Ehlmann et al., 2008; Morris et al., 2010), but the amount is limited and cannot account for the fate of a possibly denser CO_2 atmosphere on early Mars.

10.11.4.2.7 Opaline silica

Opaline silica ($SiO_2 \cdot nH_2O$) occurs at several locations on Mars (Milliken et al., 2008; Skok et al., 2010) and may have been precipitated in hydrothermal or lacustrine evaporitic environments. On Earth, it rarely persists more than a few million years. The persistence of opaline silica on Mars over much longer geologic timescales (billions of years) therefore may indicate that water was not present for extended periods of time after its original formation (Tosca and Knoll, 2009).

10.11.4.2.8 Chlorides

Chlorides have been identified in many locations in the southern highlands of Mars (Osterloo et al., 2008). The most likely formation process is via ponding and evaporation of surface water from runoff or groundwater upwelling (Osterloo et al., 2010; Ruesch et al., 2012).

10.11.4.2.9 Zeolites

Zeolites are common alteration products of basaltic ashes and palagonitic material on Earth, and they have been proposed as potential water-containing minerals on Mars. Spectral evidence for zeolites in the Martian dust has been presented by Ruff (2004) and possibly by Christensen et al. (2004b) in Mini-TES data taken in Gusev crater. Zeolites can easily be hydrated and dehydrated, and experimental studies have suggested that zeolites, if present on Mars, could undergo a diurnal cycle of hydration and dehydration with possible impact on the atmosphere (Bish et al., 2003), but Tokano and Bish (2005) showed that the diurnal atmospheric water cycle on Mars is unlikely to be affected by zeolites.

10.11.4.3 Geophysical Consequences

Based on experiments performed on dry peridotite systems, water is capable of lowering the melting temperature of mantle material by as much as 500 K. The resulting melting curve for the water-saturated system has a negative *P*–*T* gradient at low pressures, which becomes positive with increasing depth and therefore pressure. An amount as small as 0.1% of water would enable some partial melt to be formed, even if it would be impossible in the dry system. The amount of partial melting increases with increasing depth (Cattermole, 1989). Therefore, the effect of water on the solidus behavior can have a significant influence on the mantle of terrestrial planets. This is discussed in greater details in **Chapter 10.09**.

Small amounts of dissolved volatiles can have important effects on the density, viscosity, and crystallization of melts and magmas. Changing these properties influences the mechanical behavior of magmas, the differentiation, and the eruption style. The solubility of water in magmas depends on parameters like temperature, pressure, the presence of other volatiles (Williams and McBirney, 1979), and the presence of exchangeable cations, that is, those that are not in tetrahedral coordination sites (Burnham, 1979). A decrease in temperature and an increase in pressure both tend to increase the solubility of water in magma, and an increasing abundance of other volatiles decreases it (Cas and Wright, 1988). Consequently, the water contents of magmas vary widely (**Table 1**). In general, water is more soluble in highly differentiated magmas than in basaltic ones. For example, magmas from subduction zones are water-rich, because sediments and secondary altered basalts provide water sources. It seems possible that a significant fraction of the water in magmas is added in crustal magma chambers (Kyser and O'Neil, 1984). It is, however, difficult to determine the original content of H_2O, in particular the amount of water injected into the stratosphere (Schmincke, 2000).

Mierdel et al. (2007) showed recently that the asthenosphere of Earth coincides with a zone where the water solubility in mantle minerals has a pronounced minimum. The minimum is due to a sharp decrease of water solubility in aluminous orthopyroxene with depth, whereas the water solubility in olivine continuously increases with pressure. Melting in the asthenosphere may therefore be related not to volatile enrichment but to a minimum in water solubility, which causes excess water to form a hydrous silicate melt.

Dissolved water can change the density of magma and thereby influence, for example, the buoyancy. In an experiment with H_2O dissolved in a melt of albite composition ($NaAlSi_3O_8$), a partial molar volume of 14–20% $cm^3\,mol^{-1}$ has been measured for a temperature range from 750 to 950 °C. Indirect estimates for the partial molar volume of

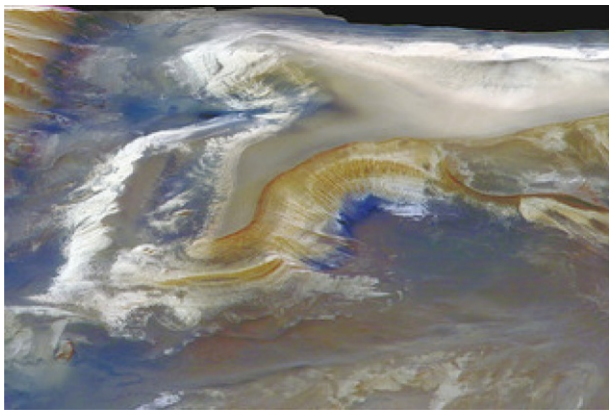


Figure 18 Layered deposit in Valles Marineris, Mars. The flank surfaces of this mountain were found to contain sulfates, based on spectra taken with the OMEGA instrument (Gendrin et al., 2005). The sulfates seem to be associated with bright layers of unknown origin. The image is a false-color perspective view, based on HRSC stereo data (image width is about 70 km, and the height of the mountain is more than 4 km). copyright: ESA/DLR/FU Berlin.

Table 1 Approximate H_2O contents in magmas

Magma content	*H_2O content (wt%)*
MORB (tholeiites) Schmincke (2000)	0.1–0.2
Island tholeiites	0.3–0.6
Alkali basalts	0.8–1.5
Subduction-zone basalts	2–3
Basanites and nephelinites	1.5–2
Andesites and dazites (island arcs)	1–2
Andesites and dazites (active continental margins)	2–4
Rhyolites	Up to ~7

MORB, mid-ocean ridge basalts.

Source: Fisher RV and Schmincke H-U (1984) *Pyroclastic Rocks*, 472 pp. Berlin, Germany: Springer; Aoki K, Ishiwaka K, and Kanisawa S (1981) Fluorine geochemistry of basaltic rocks from continental and oceanic regions and petrogenetic application. *Contributions to Mineralogy and Petrology* 76: 53–59.

water in rhyolitic or basaltic melts based on solubility data are generally at the lower end of the measurements for the albite melt (Hamilton and Oxtoby, 1986). Using the molecular weight of 18 g mol^{-1} for H_2O with the partial molar volume of $\sim$14 cm^{3}mol^{-1}, the addition of dissolved water generates a component with a density of slightly >1 g cm^{-3} compared to $\sim$2.7 and $\sim$2.3 g cm^{-3} for the anhydrous basaltic and rhyolitic melt, respectively (Wallace and Anderson, 2000). The addition of water decreases effectively the density of the melt. As an example, adding 6% dissolved water to a rhyolitic melt decreases the density by about 5%.

The viscous behavior of melts is also discussed in **Chapter 10.09**. In general, silicate melts exhibit Newtonian behavior, which means the rate of shear in the liquid is linearly proportional to the magnitude of applied shear stress. The viscosity increases rapidly with decreasing temperature. The viscosity is strongly dependent on the melt composition; with increasing SiO_2 content, the viscosity also increases. The difference in viscosity between a low-temperature rhyolitic melt and a high-temperature basaltic melt can be as much as eight orders of magnitude.

Adding dissolved water to a melt decreases the viscosity (e.g., Murase, 1962). The most likely cause is H_2O weakening or breaking apart (depolymerizing) the Si=O=Si bridges in the aluminosilicate framework of the melt (Wallace and Anderson, 2000). H_2O has a nonlinear effect on melt viscosity (Hess and Dingwell, 1996). A proportion of 0–3% of water strongly reduces the viscosity. Adding 3 wt% of water to a dry siliceous melt can decrease the viscosity by five orders of magnitude, while adding more than 3% of water has very little additional effect on the viscosity. For $<3\%$ H_2O, the water can dissolve in OH^- groups that bond to the aluminosilicate framework, thereby decreasing the viscosity very effectively. For higher amounts of water, the saturation for OH^- groups is reached and additional water is dissolved as H_2O, having little effect on the aluminosilicate bonds; therefore, little additional change in the viscosity is observed.

Water is the most important volatile species controlling the acceleration and mass eruption rate of eruption columns. Pyroclastic explosive eruptions are mainly caused by the volumetric expansion of magmatic H_2O and the resulting fragmentation of magma. This effect is substantial, as an example given by Sparks et al. (1997) makes clear: a volume of 1 m^3 rhyolitic magma at 900 °C and a H_2O content of 5 wt% will expand to a 670 m^3 of degassed mixture of water vapor and magma at 1 atm (1000 mbar) surface pressure. For small masses of up to 10–15 wt% added surface water, the formation of an eruption column in an explosive event will be promoted (Koyaguchi and Woods, 1996). On the other hand, water contents that exceed 20–30 wt% result in a dense, wet, and cold mixture, which requires a greater initial velocity to produce a buoyant column (Woods, 1995).

10.11.5 Evolution of Water and Climate

Of the terrestrial planets besides Earth, only Venus and Mars might have experienced in the past episodes in which a water-rich climate might have been possible.

Venus is at present an extremely dry and hot planet. The atmospheric water content is $\sim$30 ppmv (e.g., see de Bergh et al. (1995) and the H_2O measurements tabulated in Fegley et al. (1997)) yielding a water partial pressure of $\sim$3 mbar at the surface of Venus. For comparison, the Earth's average surface temperature is 288 K and its troposphere is much wetter with a water vapor content up to $\sim$4% (see Chapter 1.19 in Volume 1 of the *Treatise on Geochemistry* by Fegley). There are indications suggesting that Venus had more water in its past. This is based on the high deuterium/hydrogen (D/H) ratio of $(2.5\pm0.5)\times10^{-2}$ in atmospheric water vapor reported by Donahue et al. (1997). It is 160 ± 32 times higher than the D/H ratio of 1.56 ± 10^{-4} in standard mean ocean water on Earth (Lodders and Fegley, 1998). One possible scenario to explain this ratio is enrichment by isotope fractionation. Hydrogen preferentially escapes through physical processes such as evaporation or diffusion, thus enriching the deuterium concentration (e.g., see Criss, 1999). The study of the D/H ratio by Donahue et al. (1982) (Donahue et al., 1997, and references therein) concluded that Venus once must have had the equivalent of at least a 4 m deep global ocean and may have had as much as a 530 m deep global ocean. This interpretation of the D/H ratio is complicated by the fact that some interplanetary dust particles, meteorites, and comets also have a higher D/H ratio (Irvine et al., 2000). Therefore, Venus might have accreted from material with a different D/H ratio than Earth. Further clues might be derived in the future by studying the surface mineralogy. While the extreme volcanic activity of Venus has most likely erased all morphological evidence for a water-rich past, there might be traces of hydrate minerals left as indications for a water-rich past (Johnson and Fegley, 2000).

If the D/H ratio is due to a water-rich past of the planet, Venus must have lost a large amount of water during its history. Rabette et al. (2002) proposed for Earth a scenario in which an increase in sea surface temperatures above 27 °C leads to strong evaporation loading the atmosphere with a critical amount of water vapor. If enough water vapor enters the troposphere, a runaway greenhouse effect is initiated, causing even more water loss through evaporation from the ocean.

The ASPERA-4 instrument on Venus Express is studying the current oxygen and hydrogen escape rate from Venus that allows providing constraints on such escape models. Recent studies by Lammer et al. (2006) indicated that on Venus, the most relevant atmospheric escape processes of oxygen involve ions and are caused by the interaction with the solar wind. For the loss of hot H* atoms produced by photochemical reactions in the Venusian atmosphere, Gröller et al. (2010) found an average loss rate of about 3.8×10^{25}s^{-1} This is about two times lower than the estimated H^+-ion outflow from Venus' nightside due to acceleration upward along magnetic field lines of ionospheric holes (Hartle and Grebowsky, 1993) in the order of about 7×10^{25}s^{-1}. The loss rate for solar wind pickup of H^+ and O^+ ions is about 1×10^{25} and 1.6×10^{25}s^{-1}, respectively. Pickup of H_2^+ and He^+ ions is negligible compared to the loss rates of atomic hydrogen and oxygen ions. Additional loss rates are found from detached plasma clouds that yield a loss of about 5×10^{24}–1×10^{25}s^{-1} O^+ ions, in agreement with the results of a two-dimensional global hybrid simulation of Terada et al. (2002) indicating that Venus loses O^+ ions in the order of about $10^{25}\times O^+$ due to the Kelvin–Helmholtz

instability. Heavy O atoms can only escape as neutrals by atmospheric sputtering in an order of about $6 \times 10^{24} s^{-1}$ (Luhmann and Kozyra, 1991). The obtained results indicate that the ratio between H/O escape and space from the Venusian upper atmosphere is about 4 and is in a much better agreement with the stoichiometrically H/O escape ratio of 2:1, which is not the case on Mars. The ongoing detailed analysis of the outflow of ions from the Venus upper atmosphere by the ASPERA-4 and VEX-MAG instruments aboard Venus Express will lead to more accurate atmospheric loss estimations and a better understanding of the planet's water inventory.

For Mars, there are numerous studies on the atmospheric loss of water during the evolution of the planet. Chassefierre and Leblanc (2004) gave a summary of these studies and the resulting total loss of water from Mars. Zhang et al. (1993) and Luhmann et al. (1992) suggested an approach to evaluate the total amount of atmosphere lost to space along the Martian history (which is from 3.7 Gy ago up to present time) based on knowledge of what should have been the early time of the Sun. The solar extreme UV (XUV) flux was roughly three and six times higher than present XUV solar flux, 2.5 and 3.5 Gy ago, respectively (Ayres, 1997; Zahnle and Walker, 1982). This results in a globally hotter Martian thermosphere 3.5 Gy ago. Recent new model of the solar wind history (Wood et al., 2002) suggested a 40 times more massive solar wind in the early solar system (3.5 Gy ago) compared to the widely used model of Newkirk (1981). This has not been included in most models of the atmospheric loss so far. If this later process is neglected, the integrated loss to space associated with the estimates of Luhmann et al. (1992) for dissociative recombination and for pickup O^+ corona, and with the estimates of Leblanc and Johnson (2001) and Leblanc and Johnson (2002) for the sputtering, leads to a total loss to space of a H_2O ocean with a depth of 15 m covering the whole Mars surface. Lammer et al. (2003a) and Lammer et al. (2003b), using the same estimates, found a depth of 12 m. Previous estimates by Kass and Yung (1996) were of 50 m, and by Luhmann et al. (1992) and Luhmann (1997) of 30 m. Krasnopolsky and Feldman (2001) estimated the total loss of water to space based on D/H analysis and chemical model as being more than 30 m in depth. All these values have to be compared to the most recent estimate of a 150 m depth ocean of water covering the Martian surface during the Hesperian (from 3.5 Gy ago) based on geomorphological analysis (see **Section 10.11.3.4**). Carr and Head (2003) concluded that a large majority of this water should have been trapped into the surface since only the equivalent of 20–30 m depth ocean is contained in the present polar caps.

A loss of volatiles from the Martian atmosphere and a corresponding climate change are consistent with the evolution of the surface, which is recorded in the different ages of the various landforms. Fluvial and lacustrine landforms are common in ancient terrains (older than 3.5 Ga), whereas younger areas on Mars show only sparse evidence of a warmer climate that would allow the formation of such features. The morphological evidence for a climate change billions of years ago is complementary to mineralogical evidence for a more water-rich early Mars. In addition to the atmospheric data presented in the paragraph mentioned earlier, the particular clues for this climate change have been presented in the sections on morphology and mineralogy. A detailed discussion of this major climate change, the nature and timing of which is still under debate, is beyond the scope of this chapter, and the reader is referred to excellent studies by, for example, Pollack et al. (1987), Kasting (1991), Pollack (1991), Fanale et al. (1992), Jakosky and Jones (1997), Jakosky and Phillips (2001), and Bogard et al. (2001). In the following paragraph, we briefly summarize the discussion on possible, much more recent climate changes.

For Mars already early studies on Mariner 9 and Viking images showed glacial and permafrost landforms. The higher-resolution data of recent missions like MGS, Mars Odyssey, and Mars Express provided a surprising wealth of data showing details of many cold-climate phenomena. Many of them appear to be young and display a latitude-dependent, nonrandom distribution on a global scale. Such a distribution is typical for climatically controlled surface features. Although young, many of these landforms require the existence of ice and/or water at latitudes where it is not physically stable today. Moreover, morphological structures like the midlatitude cover deposit seem to be in a state of degradation. These observations suggest that recent climate excursions triggered precipitation in midlatitudes and perhaps even in low latitudes, probably as snow. One plausible reason for such climate changes in recent time could be the variable obliquity of Mars' rotational axis (Armstrong et al., 2004; Carr, 1982; Jakosky et al., 1995; Laskar and Robutel, 1993; Laskar et al., 2004; Touma and Wisdom, 1993; Ward, 1973, 1974). In contrast to Earth, whose obliquity is stabilized by the Moon (Laskar et al., 1993), the obliquity of Mars is characterized by chaotic behavior and has a high probability to reach maximum values of 60–80° and minimum values of ~0° (Laskar et al., 2004). At periods of high obliquity, the mean annual insolation at the polar regions is enhanced, and more volatiles tend to sublimate. Therefore, the polar regions experience a net loss of water, which is transported equatorward and precipitated as snow in midlatitudes and low latitudes (Levrard et al., 2004). This snow could have mantled the surface, producing a 'smoothed' appearance of the topography. Where local topographic conditions favored enhanced precipitation, larger amounts of snow could have accumulated and formed glaciers (Forget et al., 2006). In periods of low obliquity, the mean insolation at midlatitudes and low latitudes is enhanced and the ice-related landforms are degraded.

10.11.6 Summary and Outlook

Water is a highly vital, flexible, changeable, and moveable substance that alters and evolves other materials. The H_2O molecule is a major component of planets, moons, asteroids, and comets and acts in atmospheres, on surfaces, and in the subsurface as a weathering agent and transportation and sedimentation medium. It can also influence the rheology of materials and thereby influences the mantle transport processes (see **Chapter 10.09**).

Water is the main ingredient for life. As a polar molecule, it is an ideal solvent and supports several important processes for sustaining life. As water is crucial for habitable environments, its physical conditions constrain habitable zones (see **Chapter 10.14**).

Remote-sensing data show abundant evidence of fluvial and lacustrine processes that have been active on early Mars on a global scale (Noachian) with a declining rate of activity and processes limited to regional and local scales (Hesperian). These landforms also imply a large past reservoir of water on the surface and a past climate on Mars that allowed liquid water to be stable on the surface for longer timescales. Today, Mars is a cold and dry planet, but glacial, periglacial, and fluvial landforms indicate a geologically recent occurrence of water ice in the near-surface regolith, on the surface, and at least episodically in the liquid phase on the surface. The formation of these young landforms on Mars might be due to recent climate changes caused by, e.g., the variable obliquity of Mars' rotational axis.

Strictly even on Mars, all evidence for water is indirect, mainly based on remote-sensing observation of either the O–H bond (Bibring et al., 2005; Christensen et al., 2001a; Formisano et al., 2005) or the hydrogen atom (Feldman et al., 2005). The MER added morphological and indirect geochemical evidence, based on water-related alteration processes. Only in the polar areas we have direct observations of spectral features of water ice (Bibring et al., 2005; Hansen et al., 2005). The NASA Phoenix lander provided the first direct in situ observation of sublimating subsurface water ice exposed in a trench made by the robotic arm (Arvidson et al., 2009; Smith et al., 2009). Radar systems as, currently, MARSIS on Mars Express (Picardi et al., 2005) and SHARAD on MRO allow for the first time to study the subsurface structure and identify potential water or ice deposits below the surface. In combination with the data from the neutron and GRS, this allows a more accurate global inventory of water on Mars today.

Future advances in remote sensing will allow a much lower detection limit and a higher spatial resolution for mapping of water-related features. The HiRISE and CRISM instruments on MRO allow studying the morphological and mineralogical traces of water activity at a very high spatial resolution. Therefore, it will be possible to trace the history of water on Mars with a much higher confidence. The first in situ measurement of water was provided by the mass spectrometer on the Phoenix lander in the polar region of Mars (Smith et al., 2009). To further the understanding of the current-day water cycle on Mars, it is necessary to probe more hypothesized or remotely sensed water reservoirs directly. This includes direct measurements of the soil humidity at different landing sites, lander in the polar areas probing the ice and directly analyzing its composition, and access to the subsurface to probe for subsurface ice deposits. Highly mobile platforms as the NASA rover MSL or the future ESA rover ExoMars, equipped with drilling systems, neutron spectrometer, and GPR, open a new dimension in the analysis of present and past water on Mars.

The NASA MESSENGER mission has proved the existence of ice deposits close to the poles of Mercury and provided an estimate on their extent. The analysis of ice deposits on Mercury will be continued with the ESA mission BepiColombo (Benkhoff et al., 2010), as MESSENGER BepiColombo is equipped with gamma-ray and neutron spectrometers that will allow studying the spatial distribution and abundance of hydrogen in the regolith. Furthermore, the spectrometer on BepiColombo covering the wavelength range from UV to thermal IR will continue the systematic search for ice and other components of the radar-bright material. In combination with significant improvements in the knowledge of topography, the synergetic use of data provided by both spacecraft will allow to better understand the evolution and origin of water ice in the shadowed regions at the poles of Mercury.

For the Moon, the LCROSS mission confirmed the existence of water ice in permanently shadowed craters close to the poles is less certain. For the future (human) exploration of the Moon, these potential deposits are of great importance. Therefore, this will be a prime scientific priority for the new exploration program of the Moon. Further mission plans include landing and sample returns from these areas.

References

Aharonson O, Zuber MT, Neumann GA, and Head JW (1998) Mars: Northern hemisphere slopes and slope distributions. *Geophysical Research Letters* 25: 4413–4416.

Albee AL, Arvidson RE, Palluconi F, and Thorpe T (2001) Overview of the Mars Global Surveyor mission. *Journal of Geophysical Research* 106: 23291–23316.

Albee AL, Palluconi FD, and Arvidson RE (1998) Mars Global Surveyor mission: Overview and status. *Science* 279: 1671–1672.

Allen CC (1979) Areal distribution of Martian rampart craters. *Icarus* 39: 111–123.

Andreas EL (2007) New estimates for the sublimation rate for ice on the Moon. *Icarus* 186(1): 24–30.

Ansan V and Mangold N (2006) New observations of Warrego Valles, Mars: Evidence for precipitation and surface runoff. *Planetary and Space Science* 54: 219–242.

Ansan V and Mangold N (2013) 3D morphometry of valley networks on Mars from HRSC/MEX DEMs: Implications for climatic evolution through time. *Journal of Geophysical Research, Planets* 118: 1873–1894. http://dx.doi.org/10.1002/jgre.20117.

Arfstrom J and Hartmann W (2005) Martian flow features, moraine-like ridges, and gullies: Terrestrial analogs and interrelationships. *Icarus* 174: 321–335.

Armstrong JC, Leovy CB, and Quinn T (2004) A 1 Gyr climate model for Mars: New orbital statistics and the importance of seasonally resolved polar processes. *Icarus* 171: 255–271.

Arnold JR (1979) Ice in the lunar polar regions. *Journal of Geophysical Research* 84: 5659–5668.

Arvidson RE, Bonitz RG, Robinson ML, et al. (2009) Results from the Mars Phoenix Lander Robotic Arm experiment. *Journal of Geophysical Research* 114: 2156–2202.

Aston AH, Conway SJ, and Balme MR (2011) Identifying Martian gully evolution. *Geological Society, London, Special Publications* 356: 151–169.

Ayres TR (1997) Evolution of the solar ionizing flux. *Journal of Geophysical Research* 102: 1641–1651.

Baker VR (1978) The Spokane flood controversy and the Martian outflow channels. *Science* 202: 1249–1256.

Baker VR (1982) *The Channels of Mars*. Austin, TX: University of Texas Press, 198 pp.

Baker VR (2001) Water and the Martian landscape. *Nature* 412: 228–236.

Baker VR (2009) The Channeled Scabland: A retrospective. *Annual Review of Earth and Planetary Sciences* 37: 393–411. http://dx.doi.org/10.1146/annurev.earth.061008.134726.

Baker VR, Benito G, and Rudoy AN (1993) Paleohydrology of late Pleistocene superflooding, Altay Mountains, Siberia. *Science* 259: 348–350.

Baker VR, Carr MH, Gulick VC, Williams CR, and Marley MS (1992) Channels and valley networks. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 493–522. Tucson, AZ: University of Arizona Press.

Baker VR and Kochel RC (1979) Martian channel morphology – Maja and Kasei Valles. *Journal of Geophysical Research* 84: 7961–7983.

Baker VR and Milton DJ (1974) Erosion by catastrophic floods on Mars and Earth. *Icarus* 23: 27–41.

Baker VR and Nummedal D (1978) *NASA Planetary Geology Program: The Channeled Scabland*. Washington, DC: NASA.

Baker VR, Strom RG, Gulick VC, Kargel JS, and Komatsu G (1991) Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature* 352: 589–594.

Ballou EV, Wood PC, Wydeven T, Lehwalt ME, and Mack R (1978) Chemical interpretation of Viking Lander 1 life detection experiment. *Nature* 271: 644–645.

Balme M, Mangold N, Baratoux D, et al. (2006) Orientation and distribution of recent gullies in the southern hemisphere of Mars: Observations from high resolution stereo camera/Mars Express (HRSC/MEX) and Mars Orbiter camera/Mars Global

Surveyor (MOC/MGS) data. *Journal of Geophysical Research* 111: E05001. http://dx.doi.org/10.1029/2005JE002607.
Bandfield JL, Glotch TD, and Christensen PR (2003) Spectroscopic identification of carbonate minerals in the Martian dust. *Science* 301: 1084–1087.
Banerdt WB and Vidal A (2001) Surface drainage on Mars. *Lunar and Planetary Science* XXXII: 1488 (CD-ROM).
Barlow NG (2004) Martian subsurface volatile concentrations as a function of time: Clues from layered ejecta craters. *Geophysical Research Letters* 31: L05703. http://dx.doi.org/10.1029/2003GL019075.
Barlow NG (2005) A review of Martian impact crater ejecta structures and their implications for target properties. In: Kenkmann T, Hörz F, and Deutsch A (eds.) *Large Meteorite Impacts III*. *Geological Society of America Special Paper: Large Meteorite Impacts III*, vol. 384, pp. 433–442. Boulder, CO: GSA Publication.
Barlow NG, Boyce JM, Costard FM, et al. (2000) Standardizing the nomenclature of Martian impact crater ejecta morphologies. *Journal of Geophysical Research* 105: 26733–26738.
Barlow NG and Bradley TL (1990) Martian impact craters – Correlations of ejecta and interior morphologies with diameter, latitude, and terrain. *Icarus* 87: 156–179.
Barlow NG, Koroshetz J, and Dohm JM (2001) Variations in the onset diameter for Martian layered ejecta morphologies and their implications for subsurface volatile reservoirs. *Geophysical Research Letters* 28: 3095–3098.
Barnhart CJ, Howard AD, and Moore JM (2009) Long-term precipitation and late-stage valley network formation: Landform simulations of Parana Basin, Mars. *Journal of Geophysical Research* 114: E01003. http://dx.doi.org/10.1029/2008JE003122.
Barnouin-Jha OS and Schultz PH (1998) Lobateness of impact ejecta deposits from atmospheric interactions. *Journal of Geophysical Research* 103: 25739–25756.
Barsch D (1996) *Rock Glaciers*. Berlin, Germany: Springer, 331 pp.
Bart GD (2007) Comparison of small lunar landslides and Martian gullies. *Icarus* 187: 417–421.
Basilevsky AT, Werner SC, Neukum G, et al. (2006) Geologically recent tectonic, volcanic and fluvial activity on the eastern flank of the Olympus Mons volcano, Mars. *Geophysical Research Letters* 33: 1944–8007. http://dx.doi.org/10.1029/2006GL026396.
Bell JF (1996) Iron, sulfate, carbonate, and hydrated minerals on Mars. In: Burns RG, Dyar MD, McCammon C, and Schaefer MW (eds.) *Mineral Spectroscopy: A Tribute of Roger G. Burns*. *Geochemical Society Special Publication*, vol. 5, pp. 359–380. London: Geochemical Society.
Bell JF (ed.) (2008) *The Martian Surface*, p. 636. Cambridge: Cambridge University Press.
Bell JF, Squyres SW, Herkenhoff KE, et al. (2003) Mars Exploration Rover Athena panoramic camera (Pancam) investigation. *Journal of Geophysical Research* 108: 8063. http://dx.doi.org/10.1029/2003JE002070.
Benison KC and Bowen BB (2006) Acid saline lake systems give clues about past environments and the search for life on Mars. *Icarus* 183: 225–229.
Benkhoff J, van Casteren J, Hayakawa H, et al. (2010) BepiColombo—Comprehensive exploration of Mercury: Mission overview and science goals. *Planetary and Space Science* 58: 2–20.
Benn DI and Evans DJA (1998) *Glaciers and Glaciation*. London: Arnold, 734 pp.
Bennett MR and Glasser NF (1996) *Glacial Geology: Ice Sheets and Landforms*. Chichester: Wiley, 364 pp.
Berman DC, Hartmann WK, and Crown DA (2003) Debris aprons, channels, and volcanoes in the Reull Vallis region of Mars. *Lunar and Planetary Science* XXXIV: 1879.
Berman D, Hartmann W, Crown D, and Baker V (2005) The role of arcuate ridges and gullies in the degradation of craters in the Newton Basin region of Mars. *Icarus* 178: 465–486.
Bertini HW (1969) Intranuclear-cascade calculation of the secondary nucleon spectra from nucleon–nucleus interactions in the energy range 340 to 2900 MeV and comparisons with experiment. *Physical Review* 188: 1711–1730.
Bhattacharya JP, Payenberg THD, Lang SC, and Bourke M (2005) Dynamic river channels suggest a long-lived Noachian crater lake on Mars. *Geophysical Research Letters* 32: L10201. http://dx.doi.org/10.1029/2005GL022747.
Bibring J-P and Erard S (2001) The Martian surface composition. *Space Science Reviews* 96: 293–316.
Bibring J-P, Langevin Y, Gendrin A, et al. (2005) Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science* 307: 1576–1581.
Bibring J-P, Langevin Y, Mustard JF, et al. (2006) Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. *Science* 312: 400–404.
Bibring J, Langevin Y, Poulet F, et al. (2004a) Perennial water ice identified in the south polar cap of Mars. *Nature* 428: 627–630.
Bibring J-P, Langevin Y, Soufflot A, Combes C, and Cara C (1989) Results from the ISM experiment. *Nature* 341: 591–593.
Bibring J-P, Soufflot A, Berthé M, et al. (2004b) OMEGA: Observatoire pour la minèralogie, l'eau, les glaces et l'activit'e. *European Space Agency, [Special Publication]* 1240: 37–49.
Bish DL, Carey JW, Vaniman DT, and Chipera SJ (2003) Stability of hydrous minerals on the Martian surface. *Icarus* 164: 96–103.
Bishop JL (2005) Hydrated minerals on Mars. In: Tokano T (ed.) *Water on Mars and Life*, pp. 65–96. Berlin, Germany: Springer.
Bishop JL, Pieters CM, Burns RG, Edwards JO, Mancinelli RL, and Froeschl H (1995) Reflectance spectroscopy of ferric sulfate-bearing montmorillonites as Mars soil analog materials. *Icarus* 117: 101–119.
Blaney DL and McCord TB (1995) Indications of sulphate minerals in the Martian soil from Earth-based spectroscopy. *Journal of Geophysical Research* 100: 14433–14442.
Bleamaster LF and Crown DA (2005) Mantle and gully associations along the walls of Dao and Harmakhis Valles, Mars. *Geophysical Research Letters* 32: L20203. http://dx.doi.org/10.1029/2005GL023548.
Bloom AL (1991) *Geomorphology*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall, 532 pp.
Bodnar RJ (2001) PTX Phase equilibria in the H_2O–CO_2–salt system at Mars near-surface conditions. *Lunar and Planetary Science* XXXII: 1689 (CD-ROM).
Bogard DD, Clayton RN, Marti K, Owen T, and Turner G (2001) Martian volatiles: Isotopic composition, origin, and evolution. *Space Science Reviews* 96: 425–458.
Bogard DD and Johnson P (1983) Martian gases in an Antarctic meteorite. *Science* 221: 651–654.
Borg LE, Connelly JN, Nyquist LE, Shih C-Y, Wiesmann H, and Reese Y (1999) The age of the carbonates in Martian meteorite ALH84001. *Science* 286: 90–94. http://dx.doi.org/10.1126/science.286.5437.90.
Boyce JM, Mouginis-Mark P, and Garbeil H (2005) Ancient oceans in the northern lowlands of Mars: Evidence from impact crater depth/diameter relationships. *Journal of Geophysical Research* 110: 3008. http://dx.doi.org/10.1029/2004JE002328.
Boyce JM and Witbeck NE (1980) Distribution of thermal gradient values in the equatorial region of Mars based on impact crater morphology. *Reports of Planetary Geology Program, NASA TM-82385*, pp. 140–143.
Boynton WV, Feldman WC, Squyres SW, et al. (2002) Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* 297: 81–85.
Brakenridge GR (1993) Modern shelf ice, equatorial Aeolis quadrangle, Mars. In: *Lunar and Planetary Institute Conference*, pp. 175–176.
Brass GW (1980) Stability of brines on Mars. *Icarus* 42: 20–28.
Bretz H (1923) The channeled scabland of the Columbia plateau. *Journal of Geology* 31: 617–649.
Bridges JC, Catling DC, Saxton JM, Swindle TD, Lyon IC, and Grady MM (2001) Alteration assemblages in Martian meteorites: Implications for near-surface processes. *Space Science Reviews* 96: 365–392.
Bridges NT and Lackner CN (2006) Northern hemisphere Martian gullies and mantled terrain: Implications for near-surface water migration in Mars' recent past. *Journal of Geophysical Research* 111: 9014. http://dx.doi.org/10.1029/2006JE002702.
Brunauer S, Emmett PH, and Teller E (1938) Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society* 60: 309–319.
Bun Tseung J-MW and Soare RJ (2006) Thermokarst and related landforms in western Utopia Planitia Mars: Implications for near-surface excess ice. *Lunar and Planetary Science* XXXVII: 1414 (CD-ROM).
Burnham CW (1979) The importance of volatile constituents. In: Yoder HS (ed.) *The Evolution of Igneous Rocks*, pp. 439–482. Princeton, NJ: Princeton University Press.
Burns RG (1970) *Mineralogical Applications of Crystal Field Theory*. New York: Cambridge University Press, 224 pp.
Burns RG (1987) Ferric sulfates on Mars. *Journal of Geophysical Research* 92: 570–574.
Burr DM (2003) Hydraulic modelling of Athabasca Valles, Mars. *Hydrological Sciences Journal* 48: 655–664.
Burr DM, Grier JA, McEwen AS, and Keszthelyi LP (2002a) Repeated aqueous flooding from the Cerberus Fossae: Evidence for very recently extant, deep groundwater on Mars. *Icarus* 159: 53–73.
Burr DM, McEwen AS, and Sakimoto SEH (2002b) Recent aqueous floods from the Cerberus Fossae, Mars. *Geophysical Research Letters* 29: 1013. http://dx.doi.org/10.1029/2001GL013345.
Burr DM, Soare RJ, Bun Tseung J-MW, and Emery JP (2005) Young (late Amazonian), near-surface, ground ice features near the equator, Athabasca Valles, Mars. *Icarus* 178: 56–73.
Bussey DBJ, Spudis PD, and Robinson MS (1999) Illumination conditions at the lunar South Pole. *Geophysical Research Letters* 26: 1187–1190.
Butler BJ (1997) The migration of volatiles on the surface of Mercury and the Moon. *Journal of Geophysical Research* 102: 19283–19292.

Butler BJ, Muhleman DO, and Slade MA (1993) Mercury – Full-disk radar images and the detection and stability of ice at the North Pole. *Journal of Geophysical Research* 98: 15003–15023.

Cabrol NA and Grin EA (1999) Distribution, classification, and ages of Martian impact crater lakes. *Icarus* 142: 160–172.

Cabrol NA and Grin EA (2001) The evolution of lacustrine environments on Mars: Is Mars only hydrologically dormant? *Icarus* 149: 291–328.

Cabrol NA, Grin EA, and Dawidowicz G (1996) Ma'adim Vallis revisited through new topographic data: Evidence for an Ancient Intravalley Lake. *Icarus* 123: 269–283.

Cabrol NA, Grin EA, and Dawidowicz G (1997) A model of outflow generation by hydrothermal underpressure drainage in volcano-tectonic environment, Shalbatana Vallis (Mars). *Icarus* 125: 455–464.

Cabrol NA, Grin EA, and Pollard WH (2000) Possible frost mounds in an ancient Martian lake bed. *Icarus* 145: 91–107.

Calvin WM (1997) Variation of the 3-μm absorption feature on Mars: Observations over eastern Valles Marineris by the Mariner 6 infrared spectrometer. *Journal of Geophysical Research* 102: 9097–9108.

Campbell DB, Campbell BA, Carter LM, Margot J-L, and Stacy NJS (2006) No evidence for thick deposits of ice at the lunar south pole. *Nature* 443: 835–837.

Campbell DB, Chandler JF, Ostro SJ, Pettengill GH, and Shapiro II (1978) Galilean satellites – 1976 radar results. *Icarus* 34: 254–267.

Carr MH (1974) The role of lava erosion in the formation of lunar rilles and Martian channels. *Icarus* 22: 1–23.

Carr MH (1979) Formation of Martian flood features by release of water from confined aquifers. *Journal of Geophysical Research* 84: 2995–3007.

Carr MH (1982) Periodic climate change on Mars: Review of evidence and effects on distribution of volatiles. *Icarus* 50: 129–139.

Carr MH (1983) Stability of streams and lakes on Mars. *Icarus* 56: 476–495.

Carr MH (1986) Mars: A water-rich planet? *Icarus* 68: 187–216. http://dx.doi.org/10.1016/0019-1035(86)90019-9.

Carr MH (1995) The Martian drainage system and the origin of valley networks and fretted channels. *Journal of Geophysical Research* 100: 7479–7507.

Carr MH (1996) *Water on Mars*. New York: Oxford University Press, 229 pp.

Carr MH (2001) Mars Global Surveyor observations of Martian fretted terrain. *Journal of Geophysical Research* 106: 23571–23594.

Carr MH (2002) Elevations of water-worn features on Mars: Implications for circulation of groundwater. *Journal of Geophysical Research* 107: 5131. http://dx.doi.org/10.1029/2002JE001845.

Carr MH (2012) The fluvial history of Mars. *Philosophical Transactions of the Royal Society A* 370: 2193–2215. http://dx.doi.org/10.1098/rsta.2011.0500.

Carr MH and Chuang FC (1997) Martian drainage densities. *Journal of Geophysical Research* 102: 9145–9152.

Carr MH and Clow GD (1981) Martian channels and valleys – Their characteristics, distribution, and age. *Icarus* 48: 91–117.

Carr MH, Crumpler LS, Cutts JA, Greeley R, Guest JE, and Masursky H (1977) Martian impact craters and emplacement of ejecta by surface flow. *Journal of Geophysical Research* 82: 4055–4065.

Carr MH and Head JW (2003) Oceans on Mars: An assessment of the observational evidence and possible fate. *Journal of Geophysical Research* 108: 5042. http://dx.doi.org/10.1029/2002JE001963.

Carr MH and Schaber GG (1977) Martian permafrost features. *Journal of Geophysical Research* 82: 4039–4054.

Carter J, Poulet F, Bibring J-P, et al. (2013) Hydrous minerals on Mars as seen by the CRISM and OMEGA imaging spectrometers: Updated global view. *Journal of Geophysical Research* 118: 831–858. http://dx.doi.org/10.1029/2012JE004145.

Cas RAF and Wright JV (1988) *Volcanic Successions*. London: Unwin Hyman, p. 528.

Catling DC (1999) A chemical model for evaporites on early Mars: Possible sedimentary tracers of the early climate and implications for exploration. *Journal of Geophysical Research* 104: 16453–16470.

Catling DC and Moore JM (2003) The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* 165: 277–300.

Cattermole PJ (1989) *Planetary Volcanism – A Study of Volcanic Activity in the Solar System*. Chichester: Ellis Horwood, 443 pp.

Chassefierre E and Leblanc F (2004) Mars atmospheric escape and evolution: Interaction with the solar wind. *Planetary and Space Science* 52(11): 1039–1058.

Chevrier VF and Rivera-Valentin EG (2012) Formation of recurring slope lineae by liquid brines on present-day Mars. *Geophysical Research Letters* 39: L21202. http://dx.doi.org/10.1029/2012GL054119.

Christensen PR (2003) Formation of recent Martian gullies through melting of extensive water-rich snow deposits. *Nature* 422: 45–48.

Christensen PR, Anderson DL, Chase SC, et al. (1992) Thermal emission spectrometer experiment – Mars observer mission. *Journal of Geophysical Research* 97: 7719–7734.

Christensen PR, Bandfield JL, Clark RN, et al. (2000) Detection of crystalline hematite mineralization on Mars by the thermal emission spectrometer: Evidence for near-surface water. *Journal of Geophysical Research* 105: 9623–9642.

Christensen PR, Bandfield JL, Hamilton VE, et al. (2001a) Mars Global Surveyor thermal emission spectrometer experiment: Investigation description and surface science results. *Journal of Geophysical Research* 106: 23823–23872.

Christensen PR, Mehall GL, Silverman SH, et al. (2003) Miniature thermal emission spectrometer for the Mars exploration rovers. *Journal of Geophysical Research, Planets* 108: 8064. http://dx.doi.org/10.1029/2003JE002117.

Christensen PR, Morris RV, Lane MD, Bandfield JL, and Malin MC (2001b) Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. *Journal of Geophysical Research* 106: 23873–23886.

Christensen PR, Ruff SW, Fergason RL, et al. (2004a) Initial results from the mini-TES experiment in Gusev Crater from the Spirit Rover. *Science* 305: 837–842. http://dx.doi.org/10.1126/science.1100564.

Christensen PR, Wyatt MB, Glotch TD, et al. (2004b) Mineralogy at Meridiani Planum from the mini-TES experiment on the Opportunity rover. *Science* 306: 1733–1739.

Christiansen EH (1989) Lahars in the Elysium region of Mars. *Geology* 17: 203–206.

Clark RN (1981) Water frost and ice – The near-infrared spectral reflectance 0.65–2.5 microns. *Journal of Geophysical Research* 86: 3087–3096.

Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz AN (ed.) *Remote Sensing for the Earth Sciences*. *Manual of Remote Sensing*, vol. 3, pp. 3–58. New York: Wiley.

Clark RN (2009) Detection of adsorbed water and hydroxyl on the Moon. *Science* 326: 562–564. http://dx.doi.org/10.1126/science.1178105.

Clark BC III, Castro AJ, Rowe CD, et al. (1977) The Viking X ray fluorescence experiment – Analytical methods and early results. *Journal of Geophysical Research* 82: 4577–4594.

Clark RN, Swayze GA, Singer RB, and Pollack JB (1990) High-resolution reflectance spectra of Mars in the 2.3-micron region – Evidence for the mineral scapolite. *Journal of Geophysical Research* 95: 14463–14480.

Clifford SM (1987) Polar basal melting on Mars. *Journal of Geophysical Research* 92: 9135–9152.

Clifford SM (1993) A model for the hydrologic and climatic behavior of water on Mars. *Journal of Geophysical Research* 98: 10973–11016.

Clifford SM and Parker TJ (2001) The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* 154: 40–79.

Clow GD (1987) Generation of liquid water on Mars through the melting of a dusty snowpack. *Icarus* 72: 95–127.

Colaprete A and Jakosky BM (1998) Ice flow and rock glaciers on Mars. *Journal of Geophysical Research* 103: 5897–5910.

Colaprete A, Schulz P, Heldmann J, et al. (2010) Detection of water in the LCROSS Ejecta Plume. *Science* 330: 463. http://dx.doi.org/10.1126/science.1186986.

Coleman NM (2003) Aqueous flows carved the outflow channels on Mars. *Journal of Geophysical Research* 108: 5039. http://dx.doi.org/10.1029/2002JE001940.

Coleman NM (2004) Ravi Vallis, Mars paleoflood origin and genesis of secondary chaos zones. *Lunar and Planetary Science* XXXV: 1299 (CD-ROM).

Coleman NM (2005) Martian megaflood-triggered chaos formation, revealing groundwater depth, cryosphere thickness, and crustal heat flux. *Journal of Geophysical Research* 110: E12S20. http://dx.doi.org/10.1029/2005JE002419.

Conway SJ, Balme MR, Murray JB, Towner MC, Okubo CH, and Grindrod PM (2011) The indication of Martian gully formation processes by slope–area analysis. *Geological Society, London, Special Publications* 356: 171–201.

Costard FM (1989) The spatial distribution of volatiles in the Martian hydrolithosphere. *Earth, Moon, and Planets* 45: 265–290.

Costard F, Aguirre-Puente J, Greeley R, and Makhloufi N (1999) Martian fluvial–thermal erosion: Laboratory simulation. *Journal of Geophysical Research* 104(E6): 14091–14098. http://dx.doi.org/10.1029/1999JE900020.

Costard FM and Baker VR (2001) Thermokarst landforms and processes in Ares Vallis, Mars. *Geomorphology* 37: 289–301.

Costard F, Forget F, Mangold N, and Peulvast JP (2002) Formation of recent Martian debris flows by melting of near-surface ground ice at high obliquity. *Science* 295: 110–113.

Costard FM and Kargel JS (1995) Outwash plains and thermokarst on Mars. *Icarus* 114: 93–112.

Coussot P and Meunier M (1996) Recognition, classification and mechanical description of debris flows. *Earth-Science Reviews* 40: 209–227. http://dx.doi.org/10.1016/0012-8252(95)00065-8.

Craddock RA and Howard AD (2002) The case for rainfall on a warm, wet early Mars. *Journal of Geophysical Research* 107: E5111. http://dx.doi.org/10.1029/2001JE001505.

Criss RE (1999) *Principles of Stable Isotope Distribution*. New York: Oxford University Press, 254 pp.

Cutts JA (1973) Nature and origin of Martian polar layered deposits. *Bulletin of the American Astronomical Society* 78: 4231–4249.
Cutts JA and Blasius KR (1981) Origin of Martian outflow channels – The eolian hypothesis. *Journal of Geophysical Research* 86: 5075–5102.
de Bergh C, Bezard B, Crisp D, et al. (1995) Water in the deep atmosphere of Venus from high-resolution spectra of the night side. *Advances in Space Research* 15: 79–88.
De Hon R (1992) Martian lake basins and lacustrine plains. *Earth, Moon, and Planets* 56: 95–122.
De Villiers G, Kleinhans MG, and Postma G (2013) Experimental delta formation in crater lakes and implications for interpretation of Martian deltas. *Journal of Geophysical Research* 118: 651–670. http://dx.doi.org/10.1002/jgre.20069.
Di Achille G and Hynek BM (2010) Ancient Ocean on Mars supported by global distribution of deltas and valleys. *Nature Geoscience* 3: 459–463. http://dx.doi.org/10.1038/NGEO891.
Di Achille G, Marinangeli L, Ori GG, et al. (2006a) Geological evolution of the Tyras Vallis paleolacustrine system, Mars. *Journal of Geophysical Research* 111: E04003. http://dx.doi.org/10.1029/2005JE002561.
Di Achille G, Ori G-G, and Reiss D (2007) Evidence for late Hesperian lacustrine activity in Shalbatana Vallis, Mars. *Journal of Geophysical Research* 112: E07007. http://dx.doi.org/10.1029/2006JE002858.
Di Achille G, Ori GG, Reiss D, et al. (2006b) A steep fan at Coprates Catena, Valles Marineris, Mars, as seen by HRSC data. *Geophysical Research Letters* 33: L07204. http://dx.doi.org/10.1029/2005GL025435.
Dickson JL and Head JW (2009) The formation and evolution of youthful gullies on Mars: Gullies as the late-stage phase of Mars' most recent ice age. *Icarus* 204: 63–86.
Dickson J, Head J, and Kreslavsky M (2007) Martian gullies in the southern mid-latitudes of Mars: Evidence for climate-controlled formation of young fluvial features based upon local and global topography. *Icarus* 188: 315–323.
Diniega S, Byrne S, Bridges NT, Dundas CM, and McEwen AS (2010) Seasonality of present-day Martian dune-gully activity. *Geology* 38: 1047–1050.
Diniega S, Hansen CJ, McElwaine JN, et al. (2013) A new dry hypothesis for the formation of Martian linear gullies. *Icarus* 225: 526–537.
Donahue TM, Grinspoon DH, Hartle RE, and Hodges RR (1997) Ion/neutral escape of hydrogen and deuterium: Evolution of water. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 385–414. Tucson, AZ: University of Arizona Press.
Donahue TM, Hoffman JH, Hodges RR, and Watson AJ (1982) Venus was wet – A measurement of the ratio of deuterium to hydrogen. *Science* 216: 630–633.
Drake DM, Feldman WC, and Jakosky BM (1988) Martian neutron leakage spectra. *Journal of Geophysical Research* 93: 6353–6368.
Dundas CM, Diniega S, Hansen CJ, Byrne S, and McEwen AS (2012) Seasonal activity and morphological changes in Martian gullies. *Icarus* 220: 124–143.
Dundas CM, McEwen AS, Diniega S, Byrne S, and Martinez-Alonso S (2010) New and recent gully activity on Mars as seen by HiRISE. *Geophysical Research Letters* 37: L07202.
Easterbrook DJ (1999) *Surface Processes and Landforms*, 2nd edn. New York: Prentice Hall, 546 pp.
Edgett KS and Parker TJ (1997) Water on early Mars: Possible subaqueous sedimentary deposits covering ancient cratered terrain in western Arabia and Sinus Meridiani. *Geophysical Research Letters* 24: 2897–2900.
Ehlmann BL, Berger G, Mangold N, et al. (2013) Geochemical consequences of widespread clay mineral formation in Mars' ancient crust. *Space Science Reviews* 174: 329–364. http://dx.doi.org/10.1007/s11214-012-9930-0.
Ehlmann BL, Mustard JF, Murchie SL, et al. (2008) Orbital identification of carbonate-bearing rocks on Mars. *Science* 322: 1828–1832. http://dx.doi.org/10.1126/science.1164759.
Embleton C and King CAM (1975) *Glacial Geomorphology*. London: Arnold 573 pp.
Erkeling G, Reiss D, Hiesinger H, and Jaumann R (2010) Morphologic, stratigraphic and morphometric investigations of valley networks in eastern Libya Montes, Mars: Implications for the Noachian/Hesperian climate change. *Earth and Planetary Science Letters* 294: 291–305.
Estermann I (1955) Gases at low densities. In: Rossini FD (ed.) *Thermodynamics and Physics of Matter*, pp. 736–776. Princeton, NJ: Princeton University Press.
Evans DJA (2003) *Glacial Landsystems*. London: Arnold, 532 pp.
Fairén AG, Fernéndez-Remolar D, Dohm JM, Baker VR, and Amils R (2004) Inhibition of carbonate synthesis in acidic oceans on early Mars. *Nature* 431: 423–426.
Fanale FP and Cannon WA (1971) Adsorption on the Martian regolith. *Nature* 230: 502–504.
Fanale FP and Cannon WA (1974) Exchange of absorbed H_2O and CO_2 between the regolith and atmosphere of Mars caused by changes in surface insolation. *Journal of Geophysical Research* 79: 3397–3402.
Fanale FP, Postawko SE, Pollack JB, Carr MH, and Pepin RO (1992) Mars – Epochal climate change and volatile history. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 1135–1179. Tucson, AZ: University of Arizona Press.
Farmer VC (1974) *The Infra-Red Spectra of Minerals*. London: Mineralogical Society, 539 pp.
Farmer CB, Davies DW, and Laporte DD (1976) Mars – Northern summer ice cap – Water vapor observations from Viking 2. *Science* 194: 1339–1341.
Farmer CB and Doms PE (1979) Global seasonal variation of water vapor on Mars and the implications of permafrost. *Journal of Geophysical Research* 84(13): 2881–2888.
Farrand WH, Bell JF, Johnson JR, Rice MS, and Hurowitz JA (2013) VNIR multispectral observations of rocks at Cape York, Endeavour crater, Mars by the Opportunity rover's Pancam. *Icarus* 225: 709–725. http://dx.doi.org/10.1016/j.icarus.2013.04.014.
Fassett CI, Dickson JL, Head JW, Levy JS, and Marchant DR (2010) Supraglacial and proglacial valleys on Amazonian Mars. *Icarus* 208: 86–100. http://dx.doi.org/10.1016/j.icarus.2010.02.021.
Fassett CI and Head JW III (2006) Valleys on Hecates Tholus, Mars: Origin by basal melting of summit snowpack. *Planetary and Space Science* 54: 370–378.
Fassett CI and Head JW (2008a) The timing of Martian valley network activity: Constraints from buffered crater counting. *Icarus* 195: 61–89. http://dx.doi.org/10.1016/j.icarus.2007.12.009.
Fassett CI and Head JW (2008b) Valley network-fed, open-basin lakes on Mars: Distribution and implications for Noachian surface and subsurface hydrology. *Icarus* 198: 37–56. http://dx.doi.org/10.1016/j.icarus.2008.06.016.
Fassett CI, Levy JS, Head JW, and Dickson JL (2014) Estimating the volume of non-polar ice on Mars: Geometric constraints on the volume of midlatitude debris covered glaciers. In: *45th Lunar and Planetary Science Conference,* LPI Contribution No. 1777, abstract 1600.
Fastook JL, Head JW, Marchant DR, Forget F, and Madeleine JB (2012) Early Mars climate near the Noachian–Hesperian boundary: Independent evidence for cold conditions from basal melting of the south polar ice sheet (Dorsa Argentea Formation) and implications for valley network formation. *Icarus* 219: 25–40. http://dx.doi.org/10.1016/j.icarus.2012.02.013.
Fegley B Jr., Klingelhöfer G, Lodders K, and Widemann T (1997) Geochemistry of surface atmosphere interactions on Venus. In: Bougher SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 591–636. Tucson, AZ: University of Arizona Press.
Feldman WC, Boynton WV, Jakosky BM, and Mellon MT (1993) Redistribution of subsurface neutrons caused by ground ice on Mars. *Journal of Geophysical Research* 98: 20855–20870.
Feldman WC, Boynton WV, Tokar RL, et al. (2002) Global distribution of neutrons from Mars: Results from Mars Odyssey. *Science* 297: 75–78.
Feldman WC and Drake DM (1986) A Doppler filter technique to measure the hydrogen content of planetary surfaces. *Nuclear Instruments and Methods in Physics Research* 245: 182–190.
Feldman WC, Lawrence DJ, Elphic RC, et al. (2000) Polar hydrogen deposits on the Moon. *Journal of Geophysical Research* 105: 4175–4196.
Feldman WC, Maurice S, Binder AB, Barraclough BL, Elphic RC, and Lawrence DJ (1998) Fluxes of fast and epithermal neutrons from *Lunar Prospector*: Evidence for water ice at the lunar poles. *Science* 281: 1496.
Feldman WC, Prettyman TH, Maurice S, et al. (2004) Global distribution of near-surface hydrogen on Mars. *Journal of Geophysical Research* 109: E09006. http://dx.doi.org/10.1029/2003JE002160.
Feldman WC, Prettyman TH, Maurice S, et al. (2005) Topographic control of hydrogen deposits at low latitudes to midlatitudes of Mars. *Journal of Geophysical Research* 110: E11009. http://dx.doi.org/10.1029/2005JE002452.
Fisher RV and Schmincke H-U (1984) *Pyroclastic Rocks*. Berlin, Germany: Springer, 472 pp.
Flasar FM and Goody RM (1976) Diurnal behaviour of water on Mars. *Planetary and Space Science* 24: 161–181.
Fletcher NH (1970) *The Chemical Physics of Ice*. Cambridge: Cambridge University Press, 271 pp.
Forget F, Haberle RM, Montmessin F, Levrard B, and Head JW (2006) Formation of glaciers on Mars by atmospheric precipitation at high obliquity. *Science* 311: 368–371.
Formisano V, Angrilli F, Arnold G, et al. (2005) The planetary Fourier spectrometer (PFS) onboard the European Mars Express mission. *Planetary and Space Science* 53: 963–974.
Forsythe RD and Zimbelman JR (1995) A case for ancient evaporite basins on Mars. *Journal of Geophysical Research* 100: 5553–5563.
French HM (1996) *The Periglacial Environment*, 2nd edn. Essex: Addison Wesley Longman, 341 pp.

Gaidos EJ (2001) Note: Cryovolcanism and the recent flow of liquid water on Mars. *Icarus* 153: 218–223.

Gendrin A, Mangold N, Bibring J, et al. (2005) Sulfates in Martian layered terrains: The OMEGA/Mars Express view. *Science* 307: 1587–1591.

Ghatan GJ, Head JW, and Wilson L (2005) Mangala Valles, Mars: Assessment of early stages of flooding and downstream flood evolution. *Earth, Moon, and Planets* 96: 1–57. http://dx.doi.org/10.1007/s11038-005-9009-y.

Ghatan GJ and Zimbelman JR (2006) Paucity of candidate coastal constructional landforms along proposed shorelines on Mars: Implications for a northern lowlands-filling ocean. *Icarus* 185(1): 171–196.

Gilmore MS and Phillips EL (2002) Role of aquicludes in formation of Martian gullies. *Geology* 30: 1107–1110.

Glen JW (1955) The creep of polycrystalline ice. *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 228(1175): 519–538.

Goetz W, Bertelsen P, Binau CS, et al. (2005) Indication of drier periods on Mars from the chemistry and mineralogy of atmospheric dust. *Nature* 436: 62–65.

Goldspiel JM and Squyres SW (2000) Groundwater sapping and valley formation on Mars. *Icarus* 148: 176–192.

Goldspiel JM, Squyres SW, and Jankowski DG (1993) Topography of small Martian valleys. *Icarus* 105: 479–500.

Golombek MP and Bridges NT (2000) Erosion rates on Mars and implications for climate change: Constraints from the pathfinder landing site. *Journal of Geophysical Research* 105: 1841–1854.

Golombek MP, Grant JA, Crumpler LS, et al. (2006) Erosion rates at the Mars exploration rover landing sites and long-term climate change on Mars. *Journal of Geophysical Research* 111: E12S10. http://dx.doi.org/10.1029/2006JE002754.

Golombek MP, Grant JA, Parker TJ, et al. (2003) Selection of the Mars exploration rover landing sites. *Journal of Geophysical Research* 108: 8072. http://dx.doi.org/10.1029/2003JE002074.

Goudge TA, Head JW, Mustard JF, and Fassett CI (2012) An analysis of open-basin lake deposits on Mars: Evidence for the nature of associated lacustrine deposits and post-lacustrine modification processes. *Icarus* 219: 211–229. http://dx.doi.org/10.1016/j.icarus.2012.02.027.

Grant JA (2000) Valley formation in Margaritifer Sinus, Mars, by precipitation-recharged ground-water sapping. *Geology* 28: 223–226.

Grant JA and Wilson SA (2011) Late alluvial fan formation in southern Margaritifer Terra, Mars. *Geophysical Research Letters* 38: L08201. http://dx.doi.org/10.1029/2011GL046844.

Greeley R (1985) *Planetary Landscapes*. London: Allen & Unwin, 286 pp.

Greeley R and Guest JE (1987) *US Geological Survey Miscellaneous Investigation Series, I-1802-B: Geologic Map of the Eastern Equatorial Region of Mars*. Flagstaff, AZ: US Geological Survey.

Greeley R and Spudis PD (1981) Volcanism on Mars. *Reviews of Geophysics and Space Physics* 19: 13–41.

Grimm RE (2002) Low-frequency electromagnetic exploration for groundwater on Mars. *Journal of Geophysical Research* 107: 5006. http://dx.doi.org/10.1029/2001JE001504.

Grin EA and Cabrol NA (1997) Limnologic analysis of Gusev crater paleolake, Mars. *Icarus* 130: 461–474.

Gröller H, Shematovich VI, Lichtenegger HIM, et al. (2010) Venus' atomic hot oxygen environment. *Journal of Geophysical Research* 115: E12017.

Grosswald MG (1998) New approach to the Ice Age paleohydrology of northern Eurasia. In: Benito G, Baker VR, and Gregory KJ (eds.) *Paleohydrology and Environmental Change*, pp. 199–214. New York: Wiley.

Grotzinger JP, Arvidson RE, Bell JF, et al. (2005) Stratigraphy and sedimentology of a dry to wet eolian depositional system, burns formation, Meridiani Planum, Mars. *Earth and Planetary Science Letters* 240: 11–72.

Gulick VC (2001) Origin of the valley networks on Mars: A hydrological perspective. *Geomorphology* 37: 241–268.

Gulick VC and Baker VR (1989) Fluvial valleys and Martian palaeoclimates. *Nature* 341: 514–516.

Gulick VC and Baker VR (1990) Origin and evolution of valleys on Martian volcanoes. *Journal of Geophysical Research* 95: 14325–14344.

Haberle RM, McKay CP, Schaeffer J, et al. (2001) On the possibility of liquid water on present-day Mars. *Journal of Geophysical Research* 106: 23317–23326.

Hambrey MJ (1994) *Glacial Environments*. London: UCL Press, 296 pp.

Hamilton WC and Ibers JA (1968) *Hydrogen Bonding in Solids: Methods of Molecular Structure Determination*. New York: W. A. Benjamin.

Hamilton DL and Oxtoby S (1986) Solubility of water in albite-melt determined by the weight-loss method. *Journal of Geology* 94: 626–630.

Hansen G, Giuranna M, Formisano V, et al. (2005) PFS-MEX observation of ices in the residual south polar cap of Mars. *Planetary and Space Science* 53: 1089–1095.

Hapke B and Blewett D (1991) Coherent backscatter model for the unusual radar reflectivity of icy satellites. *Nature* 352: 46–47.

Hargraves RB, Collinson DW, Arvidson RE, and Spitzer CR (1977) The Viking magnetic properties experiment – Primary mission results. *Journal of Geophysical Research* 82: 4547–4558.

Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's North Pole. *Icarus* 149: 1–15.

Harmon JK and Slade MA (1992) Radar mapping of Mercury – Full-disk images and polar anomalies. *Science* 258: 640–643.

Harrison KP and Grimm RE (2004) Tharsis recharge: A source of groundwater for Martian outflow channels. *Geophysical Research Letters* 31: 14703. http://dx.doi.org/10.1029/2004GL020502.

Harrison KP and Grimm RE (2005) Groundwater-controlled valley networks and the decline of surface runoff on early Mars. *Journal of Geophysical Research* 110: E12S16. http://dx.doi.org/10.1029/2005JE002455.

Hartle RE and Grebowsky JM (1993) Light ion flow in the nightside ionosphere of Venus. *Journal of Geophysical Research* 98: 7437–7445.

Hartmann WK (2001) Martian seeps and their relation to youthful geothermal activity. *Space Science Reviews* 96: 405–410.

Hasegawa S, Murakawa K, Ishiguro M, et al. (2003) Evidence of hydrated and/or hydroxylated minerals on the surface of asteroid 4 Vesta. *Geophysical Research Letters* 30: 2123. http://dx.doi.org/10.1029/2003GL018627.

Hasted JB (1973) *Aqueous Dielectrics*. London: Chapman and Hall, 302 pp.

Hauber E, Gwinner K, Kleinhans M, et al. (2009) Sedimentary deposits in Xanthe Terra: Implications for the ancient climate on Mars. *Planetary and Space Science* 57: 944–957. http://dx.doi.org/10.1016/j.pss.2008.06.009.

Hauber E, Platz T, Reiss D, et al. (2013) Asynchronous formation of Hesperian and Amazonian-aged deltas on Mars and implications for climate. *Journal of Geophysical Research* 118: 1529–1544. http://dx.doi.org/10.1002/jgre.20107.

Hauber E, van Gasselt S, Chapman MG, and Neukum G (2008) Geomorphic evidence for former lobate debris aprons at low latitudes on Mars: Indicators of the Martian paleoclimate. *Journal of Geophysical Research* 113: E02007. http://dx.doi.org/10.1029/2007JE002897.

Hauber E, van Gasselt S, Ivanov B, et al. (2005) Discovery of a flank caldera and very young glacial activity at Hecates Tholus, Mars. *Nature* 434: 356–361.

Head JW III, Kreslavsky M, Hiesinger H, and Pratt S (1999) Northern seas and oceans in the past history of Mars: New evidence from *Mars Orbiter* laser altimeter (MOLA) data. *Lunar and Planetary Science* XXX: 1352 (CD-ROM).

Head JW, Kreslavsky M, Hiesinger H, et al. (1998) Oceans in the past history of Mars: Tests for their presence using Mars Orbiter laser altimeter (MOLA) data. *Geophysical Research Letters* 25: 4401–4404.

Head JW, Kreslavsky MA, and Pratt S (2002) Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *Journal of Geophysical Research* 107: 5003. http://dx.doi.org/10.1029/2000JE001445.

Head JW and Marchant DR (2003) Cold-based mountain glaciers on Mars: Western Arsia Mons. *Geology* 31: 641–644.

Head JW, Marchant DR, Agnew MC, Fassett CI, and Kreslavsky MA (2006) Extensive valley glacier deposits in the northern mid-latitudes of Mars: Evidence for Late Amazonian obliquity-driven climate change. *Earth and Planetary Science Letters* 241: 663–671.

Head JW, Marchant DR, and Kreslavsky MA (2008) Formation of gullies on Mars: Link to recent climate history and insolation microenvironments implicate surface water flow origin. *Proceedings of the National Academy of Sciences of the United States of America* 105: 13258–13263.

Head JW, Mustard JF, Kreslavsky MA, Milliken RE, and Marchant DR (2003a) Recent ice ages on Mars. *Nature* 426: 797–802.

Head JW, Neukum G, Jaumann R, et al. (2005) Tropical to midlatitude snow and ice accumulation, flow and glaciation on Mars. *Nature* 434: 346–351.

Head JW and Roth R (1976) Mars pedestal crater escarpments: Evidence for ejecta-related emplacement. In: IPI Contribution No. 259: Papers Presented to the Symposium on Planetary Cratering Mechanics, pp. 50–52. Houston, TX: Lunar Science Institute.

Head JW, Wilson L, and Mitchell KL (2003b) Generation of recent massive water floods at Cerberus Fossae, Mars by dike emplacement, cryospheric cracking, and confined aquifer groundwater release. *Geophysical Research Letters* 30: 1577. http://dx.doi.org/10.1029/2003GL017135.

Heggy E, Paillou P, Ruffie G, Malezieux JM, Costard F, and Grandjean G (2001) On water detection in the Martian subsurface using sounding radar. *Icarus* 154: 244–257.

Helbert J, Reiss D, Hauber E, and Benkhoff J (2005) Limits on the burial depth of glacial ice deposits on the flanks of Hecates Tholus, Mars. *Geophysical Research Letters* 32: L17201. http://dx.doi.org/10.1029/2005GL023712.

Heldmann JL, Carlsson E, Johansson H, Mellon MT, and Toon OB (2007) Observations of Martian gullies and constraints on potential formation mechanisms: II. The northern hemisphere. *Icarus* 188: 324–344.
Heldmann JL, Conley CA, Brown AJ, Fletcher L, Bishop JL, and McKay CP (2010) Possible liquid water origin for Atacama Desert mudflow and recent gully deposits on Mars. *Icarus* 206: 685–690.
Heldmann JL and Mellon MT (2004) Observations of Martian gullies and constraints on potential formation mechanisms. *Icarus* 168: 285–304.
Herkenhoff KE and Murray BC (1990) High-resolution topography and albedo of the South Polar layered deposits on Mars. *Journal of Geophysical Research* 95: 14511–14529.
Hess KU and Dingwell DB (1996) Viscosities of hydrous leucogranitic melts: A non-Arrhenian model. *American Mineralogist* 81: 1297–1300.
Hiesinger H and Head JW (2000) Characteristics and origin of polygonal terrain in southern Utopia Planitia, Mars: Results from Mars Orbiter laser altimeter and Mars Orbiter camera data. *Journal of Geophysical Research* 105(E5): 11999–12022. http://dx.doi.org/10.1029/1999JE001193.
Hoffman N (2000a) White Mars: A new model for Mars' surface and atmosphere based on CO_2. *Icarus* 146: 326–342.
Hoffman N (2000b) Ideas about the surface runoff features on Mars. *Science* 290: 711–714.
Hoffman N (2002) Active polar gullies on Mars and the role of carbon dioxide. *Astrobiology* 2: 313–323. http://dx.doi.org/10.1089/153110702762027899.
Hoke MRT and Hynek BM (2009) Roaming zones of precipitation on ancient Mars as recorded in valley networks. *Journal of Geophysical Research* 114: E08002. http://dx.doi.org/10.1029/2008JE003247.
Hoke MRT, Hynek BM, Di Achille G, and Hutton EWH (2014) The effects of sediment supply and concentrations on the formation timescale of Martian deltas. *Icarus* 228: 1–12. http://dx.doi.org/10.1016/j.icarus.2013.09.017.
Howard AD and Kochel RC (1988) Introduction to cuesta landforms and sapping processes on the Colorado Plateau. In: Howard AD, Kochel RC, and Holt HR (eds.) *Sapping Features of the Colorado Plateau*. *NASA Special Publication*, vol. 491, pp. 6–56. Washington, DC: NASA.
Howard AD and Moore JM (2011) Late Hesperian to early Amazonian midlatitude Martian valleys: Evidence from Newton and Gorgonum basins. *Journal of Geophysical Research* 116: E05003. http://dx.doi.org/10.1029/2010JE003782.
Howard AD, Moore JM, and Irwin RP (2005) An intense terminal epoch of widespread fluvial activity on early Mars. 1: Valley network incision and associated deposits. *Journal of Geophysical Research* 110: E12S14. http://dx.doi.org/10.1029/2005JE002459.
Hugenholtz C (2008) Frosted granular flow: A new hypothesis for mass wasting in Martian gullies. *Icarus* 197: 65–72.
Huggett RJ (2003) *Fundamentals of Geomorphology*. London: Routledge, 386 pp.
Hynek BM, Beach M, and Hoke MRT (2010) Updated global map of Martian valley networks and implications for climate and hydrologic processes. *Journal of Geophysical Research* 115: E09008. http://dx.doi.org/10.1029/2009JE003548.
Hynek BM and Phillips RJ (2003) New data reveal mature, integrated drainage systems on Mars indicative of past precipitation. *Lunar and Planetary Science* XXXIV: 1842 (CDROM).
Ingersoll AP, Svitek T, and Murray BC (1992) Stability of polar frosts in spherical bowl-shaped craters on the moon, Mercury, and Mars. *Icarus* 100: 40–47.
Irvine WM, Schloerb FP, Crovisier J, Fegley B Jr., and Mumma MJ (2000) Comets: A link between interstellar and nebular chemistry. In: Boss A, Manning V, and Russell S (eds.) *Protostars and Planets IV*, p. 1159. Tucson, AZ: University of Arizona Press.
Irwin RP III, Craddock RA, and Howard AD (2005) Interior channels in Martian valley networks: Discharge and runoff production. *Geology* 33: 489–492.
Irwin RP and Howard AD (2002) Drainage basin evolution in Noachian Terra Cimmeria, Mars. *Journal of Geophysical Research* 107: 5056. http://dx.doi.org/10.1029/2001JE001818.
Irwin RP, Howard AD, and Craddock RA (2008) Fluvial valley networks on Mars. In: Rice SP, et al. (ed.) *River Confluences, Tributaries, and the Fluvial Network*, pp. 419–451. Chichester: Wiley.
Irwin RP, Howard AD, Craddock RA, and Moore JM (2005a) An intense terminal epoch of widespread fluvial activity on early Mars. 2: Increased runoff and paleolake development. *Journal of Geophysical Research* 110: E12S12. http://dx.doi.org/10.1029/2005JE002460.
Ivanov MA and Head JW (2001) Chryse Planitia, Mars: Topographic configuration, outflow channel continuity and sequence, and tests for hypothesized ancient bodies of water using Mars Orbiter laser altimeter (MOLA) data. *Journal of Geophysical Research* 106: 3275–3296.
Iverson RM (1997) The physics of debris flows. *Reviews of Geophysics* 35: 245–296. http://dx.doi.org/10.1029/97RG00426.
Jahn A (1975) *Problems of the Periglacial Zone*. Warschau: Polish Scientific Publications, 223 pp.
Jakosky BM (1985) The seasonal cycle of water on Mars. *Space Science Reviews* 41: 131–200.
Jakosky BM, Henderson BG, and Mellon MT (1995) Chaotic obliquity and the nature of the Martian climate. *Journal of Geophysical Research* 100: 1579–1584.
Jakosky BM and Jones JH (1997) The history of Martian volatiles. *Reviews of Geophysics* 35: 1–16.
Jakosky BM and Phillips RJ (2001) Mars' volatile and climate history. *Nature* 412: 237–244.
James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, AZ: University of Arizona Press.
Jänchen J, Bish DL, Möhlmann DTF, and Stach H (2006) Investigation of the water sorption properties of Mars-relevant micro- and mesoporous minerals. *Icarus* 180: 353–358.
Jankowski DG and Squyres SW (1992) The topography of impact craters in 'softened' terrain on Mars. *Icarus* 100: 26–39. http://dx.doi.org/10.1016/0019-1035(92)90015-Y.
Jerolmack DJ, Mohrig D, Zuber MT, and Byrne S (2004) A minimum time for the formation of Holden Northeast fan, Mars. *Geophysical Research Letters* 31: L21701. http://dx.doi.org/10.1029/2004GL021326.
Johnson NM and Fegley B (2000) Water on Venus: New insights from tremolite decomposition. *Icarus* 146: 301–306.
Johnson CL, Solomon SC, Head JW, Phillips RJ, Smith DE, and Zuber MT (2000) Lithospheric loading by the northern polar cap on Mars. *Icarus* 144: 313–328.
Jöns H-P (1984) Sedimentary basins and mud flows in the northern lowlands of Mars. In: *Lunar and Planetary Institute Conference*, pp. 417–418.
Jöns H-P (1985) Late sedimentation and late sediments in the northern lowlands on Mars. In: *Lunar and Planetary Institute Conference*, pp. 414–415.
Jöns H-P (1986) Arcuate ground undulations, gelifluxionlike features, and 'Front Tori' in the northern lowlands of Mars – What do they indicate? In: *Lunar and Planetary Institute Conference*, pp. 404–405.
Jöns H-P (2002) Junge/Rezente nichtäolische exogene dynamik auf dem Mars: Großflächige Beeinflussung durch zerfallende CO_2–hydrate? *Zeitschrift für Geologische Wissenschaften* 6: 403–421.
Jouannic G, Gargani J, Costard F, et al. (2012) Morphological and mechanical characterization of gullies in a periglacial environment: The case of the Russell crater dune (Mars). *Planetary and Space Science* 71: 38–54.
Jouglet D, Poulet F, Mustard JF, et al. (2006) Observation of 3 μm hydration feature on Mars from OMEGA-MEx data. *Lunar and Planetary Science* XXXVII: 1741 (CD-ROM).
Kargel JS, Baker VR, Beget JE, et al. (1995) Evidence of ancient continental glaciation in the Martian northern plains. *Journal of Geophysical Research* 100: 5351–5368.
Kargel JS and Strom RG (1992) Ancient glaciation on Mars. *Geology* 20: 3–7.
Kass DM and Yung YL (1996) The loss of atmosphere from Mars response. *Science* 274: 1932–1933.
Kasting JF (1991) CO_2 condensation and the climate of early Mars. *Icarus* 94: 1–13.
Kasting JF, Whitmire DP, and Reynolds RT (1993) Habitable zones around main sequence stars. *Icarus* 101: 108–128.
Keszthelyi L, Jaeger W, Dundas CM, Williams DA, and Cataldo V (2014) Evidence for possible mechanical erosion by lava at Athabasca Valles, Mars, from HiRISE and CTX images and topography. In: *45th Lunar and Planetary Science Conference,* LPI Contribution No. 1777, Abstract 1683.
Keszthelyi L, Thordarson T, McEwen A, et al. (2004) Icelandic analogs to Martian flood lavas. *Geochemistry, Geophysics, Geosystems* 5: Q11014.
Kieffer HH (1979) Mars south polar spring and summer temperatures – A residual CO_2 frost. *Journal of Geophysical Research* 84: 8263–8288.
Kieffer HH, Martin TZ, Chase SC, Miner ED, and Palluconi FD (1976) Martian north pole summer temperatures – Dirty water ice. *Science* 194: 1341–1344.
Kieffer HH, Titus TN, Mullins KF, and Christensen PR (2000) Mars south polar spring and summer behavior observed by TES: Seasonal cap evolution controlled by frost grain size. *Journal of Geophysical Research* 105: 9653–9700.
Kleinhans MG (2005) Flow discharge and sediment transport models for estimating a minimum timescale of hydrological activity and channel and delta formation on Mars. *Journal of Geophysical Research* 110: 12003. http://dx.doi.org/10.1029/2005JE002521.
Kleinhans MG, van de Kasteele HE, and Hauber E (2010) Palaeoflow reconstruction from fan delta morphology on Mars. *Earth and Planetary Science Letters* 294: 378–392. http://dx.doi.org/10.1016/j.epsl.2009.11.025.
Klingelhöfer G, Morris RV, Bernhardt B, et al. (2003) Athena MIMOS II Mössbauer spectrometer investigation. *Journal of Geophysical Research* 108: 8067. http://dx.doi.org/10.1029/2003JE002138.

Klingelhöfer G, Morris RV, Bernhardt B, et al. (2004) Jarosite and hematite at Meridiani Planum from Opportunity's Mössbauer spectrometer. *Science* 306: 1740–1745.

Kneissl T, Reiss D, van Gasselt S, and Neukum G (2010) Distribution and orientation of northern-hemisphere gullies on Mars from the evaluation of HRSC and MOC-NA data. *Earth and Planetary Science Letters* 294: 357–367.

Kolb KJ, Pelletier JD, and McEwen AS (2010) Modeling the formation of bright slope deposits associated with gullies in Hale Crater, Mars: Implications for recent liquid water. *Icarus* 205: 113–137.

Komar PD (1979) Comparisons of the hydraulics of water flows in Martian outflow channels with flows of similar scale on earth. *Icarus* 37: 156–181.

Komar PD (1980) Modes of sediment transport in channelized water flows with ramifications to the erosion of the Martian outflow channels. *Icarus* 42: 317–329.

Koyaguchi T and Woods AW (1996) On the formation of eruption columns following explosive mixing of magma and surface-water. *Journal of Geophysical Research* 101: 5561–5574.

Kraal ER, Asphaug E, Moore JM, et al. (2008a) Catalogue of large alluvial fans in Martian impact craters. *Icarus* 194: 101–110.

Kraal ER, van Dijk M, Postma G, and Kleinhans MG (2008b) Martian stepped-delta formation by rapid water release. *Nature* 451: 973–976. http://dx.doi.org/10.1038/nature06615.

Krasnopolsky VA and Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294: 1914–1917.

Kreslavsky MA and Head JW (1999) Kilometer-scale slopes on Mars and their correlation with geologic units: Initial results from Mars Orbiter laser altimeter (MOLA) data. *Journal of Geophysical Research* 104: 21911–21924.

Kreslavsky MA and Head JW (2000) Kilometer-scale roughness of Mars: Results from MOLA data analysis. *Journal of Geophysical Research* 105: 26695–26712.

Kreslavsky MA and Head JW (2002) Fate of outflow channel effluents in the northern lowlands of Mars: The Vastitas Borealis Formation as a sublimation residue from frozen ponded bodies of water. *Journal of Geophysical Research, Planets* 107: 5121. http://dx.doi.org/10.1029/2001JE001831.

Kuzmin RO (1980) Determination of frozen soil depth on Mars from the morphology of fresh craters. *Akademiia Nauk SSSR Doklady* 252: 1445–1448.

Kuzmin RO (2005) Ground ice in the Martian regolith. In: Tokano T (ed.) *Advances in Astrobiology and Biogeophysics: Water on Mars and Life*, pp. 155–189. Berlin, Germany: Springer.

Kuzmin RO, Bobina NN, Zabalueva EV, and Shashkina VP (1988) Structural inhomogeneities of the Martian cryolithosphere. *Solar System Research* 22: 121–133.

Kuzmin RO, Bobina NN, Zabalueva EV, and Shashkina VP (1989) Structural irregularities of the upper layers of the Martian cryolithosphere. In: *Space Chemistry and Comparative Planetology*, pp. 80–95. Dordrecht: Kluwer.

Kuzmin RO and Zabalueva EV (2003) Polygonal terrains on Mars: Preliminary results of global mapping of their spatial distribution. *Lunar and Planetary Science* XXXIV (CD-ROM).

Kuzmin RO and Zabalueva EV (1998) On salt solutions in the Martian cryolithosphere. *Astronomicheskii Vestnik* 32: 187.

Kyser TK and O'Neil JR (1984) Hydrogen isotope systematics of submarine basalts. *Geochimica et Cosmochimica Acta* 48: 2123–2133.

Lachenbruch AH (1962) Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost. *Geological Society of America, Special Paper* 70: 69.

Laity JE and Malin MC (1985) Sapping processes and the development of theatre-headed valley networks on the Colorado Plateau. *Geological Society of America Bulletin* 96: 203–217.

Lajeunesse E, Quantin C, Allemand P, and Delacourt C (2006) New insights on the runout of large landslides in the Valles-Marineris canyons, Mars. *Geophysical Research Letters* 33(4): L04403. http://dx.doi.org/10.1029/2005GL025168.

Lamb MP, Howard AD, Johnson J, Whipple KX, Dietrich WE, and Perron JT (2006) Can springs cut canyons into rock? *Journal of Geophysical Research* 111: 7002. http://dx.doi.org/10.1029/2005JE002663.

Lammer H, Chassefière E, Karatekin Ö, et al. (2013) Outgassing history and escape of the Martian atmosphere and water inventory. *Space Science Reviews* 174: 113–154. http://dx.doi.org/10.1007/s11214-012-9943-8.

Lammer H, Kolb C, Penz T, Amerstorfer UV, Biernat HK, and Bodiselitsch B (2003a) Estimation of the past and present Martian water-ice reservoirs by isotopic constraints on exchange between the atmosphere and the surface. *International Journal of Astrobiology* 2: 1–8.

Lammer H, Lichtenegger HIM, Biernat HK, et al. (2006) Loss of hydrogen and oxygen from the upper atmosphere of Venus. *Planetary and Space Science* 54(13–14): 1445–1456.

Lammer H, Lichtenegger HIM, Kolb C, et al. (2003b) Loss of water from Mars implications for the oxidation of the soil. *Icarus* 165: 9–25.

Lane MD, Dyar MD, and Bishop JL (2004) Spectroscopic evidence for hydrous iron sulfate in the Martian soil. *Geophysical Research Letters* 31: L19702. http://dx.doi.org/10.1029/2004GL021231.

Langevin Y, Poulet F, Bibring J-P, and Gondet B (2005a) Sulfates in the north polar region of Mars detected by OMEGA/Mars Express. *Science* 307: 1584–1586.

Langevin Y, Poulet F, Bibring J-P, Schmitt B, Douté S, and Gondet B (2005b) Summer evolution of the north polar cap of Mars as observed by OMEGA/Mars Express. *Science* 307: 1581–1584.

Langmuir L (1915) Chemical reactions at low pressures. *Journal of American Chemical Society* 37: 1139–1167.

Lanz J (2004) *Geometrische, morphologische und stratigraphische Untersuchungen ausgewählter Outflow Channel der Circum-Chryse-Region, Mars, mit Methoden der Fernerkundung*. Dissertation, Free University, Berlin, DLRForschungsbericht 2004-02, 180 pp.

Lanza NL, Meyer Ga, Okubo CH, Newsom HE, and Wiens RC (2010) Evidence for debris flow gully formation initiated by shallow subsurface water on Mars. *Icarus* 205: 103–112.

Laskar J, Correia ACM, Gastineau M, Joutel F, Levrard B, and Robutel P (2004) Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* 170: 343–364.

Laskar J, Joutel F, and Robutel P (1993) Stabilization of the earth's obliquity by the moon. *Nature* 361: 615–617.

Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.

Lasue J, Mangold N, Hauber E, et al. (2013a) Quantitative assessments of the Martian hydrosphere. *Space Science Reviews* 174: 155–212. http://dx.doi.org/10.1007/s11214-012-9946-5.

Lasue J, Maurice S, Meslin PY, et al. (2013b) *ChemCam Analysis of Martian Fine Dust*. London: EPSC, Abstracts 8, EPSC2013-867.

Lawrence D, Feldman W, Goldsten J, et al. (2013) Evidence for water ice near Mercury's north pole from MESSENGER Neutron Spectrometer measurements. *Science* 80: 1–6.

Le Deit L, Flahaut J, Quantin C, et al. (2012) Extensive surface pedogenic alteration of the Martian Noachian crust suggested by plateau phyllosilicates around Valles Marineris. *Journal of Geophysical Research* 117: E00J05. http://dx.doi.org/10.1029/2011JE003983.

Leask HJ, Wilson L, and Mitchell KL (2006a) Formation of Ravi Vallis outflow channel, Mars: Morphological development, water discharge, and duration estimates. *Journal of Geophysical Research* 111: 8070. http://dx.doi.org/10.1029/2005JE002550.

Leask HJ, Wilson L, and Mitchell KL (2006b) Formation of Aromatum chaos, Mars: Morphological development as a result of volcano–ice interactions. *Journal of Geophysical Research* 111: 8071. http://dx.doi.org/10.1029/2005JE002549.

Leblanc F and Johnson RE (2001) Sputtering of the Martian atmosphere by solar wind pick-up ions. *Planetary and Space Science* 49: 645–656.

Leblanc F and Johnson RE (2002) Role of molecules in pick-up ion sputtering of the Martian atmosphere. *Journal of Geophysical Research* 107: 5010, 10209/2000JE001473.

Lee P, Cockell CS, Marinova MM, McKay CP, and Rice JW (2001) Snow and ice melt flow features on Devon Island, Nunavut, Arctic Canada as possible analogs for recent slope flow features on Mars. *Lunar and Planetary Science* XXXII: 1809 (CD-ROM).

Lefort A, Russell PS, and Thomas N (2010) Scalloped terrain in Peneus and Amphitrites Paterae region of Mars as observed by HiRISE. *Icarus* 205: 259–268. http://dx.doi.org/10.1016/j.icarus.2009.06.005.

Leverington DW (2004) Volcanic rilles, streamlined islands, and the origin of outflow channels on Mars. *Journal of Geophysical Research* 109: 10011. http://dx.doi.org/10.1029/2004JE002311.

Leverington DW (2011) A volcanic origin for the outflow channels of Mars: Key evidence and major implications. *Geomorphology* 132: 51–75. http://dx.doi.org/10.1016/j.geomorph.2011.05.022.

Levrard B, Forget F, Montmessin F, and Laskar J (2004) Recent ice-rich deposits formed at high latitudes on Mars by sublimation of unstable equatorial ice during low obliquity. *Nature* 431: 1072–1075.

Levy J (2012) Hydrological characteristics of recurrent slope lineae on Mars: Evidence for liquid flow through regolith and comparisons with Antarctic terrestrial analogs. *Icarus* 219: 1–4.

Levy JS, Head JW, Dickson JL, Fassett CI, Morgan GA, and Schon SC (2010) Identification of gully debris flow deposits in Protonilus Mensae, Mars: Characterization of a water-bearing, energetic gully-forming process. *Earth and Planetary Science Letters* 294: 368–377.

Li H, Robinson MS, and Jurdy DM (2005) Origin of Martian northern hemisphere midlatitude lobate debris aprons. *Icarus* 176: 382–394.

Lodders K and Fegley BF (1998) *Treatise on Geochemistry*, vol. 1, ch. 19.

Lozinski W (1909) Über die mechanische Verwitterung der Sandsteine im gemässigten Klima. *Bulletin of International L' Academic des Sciences de Gacouie class des Sciences Mathematique et Naturalles* 1: 1–25.

Lucchitta BK (1979) Landslides in Valles Marineris, Mars. *Journal of Geophysical Research* 84: 8097–8113.

Lucchitta BK (1981) Mars and earth – Comparison of cold-climate features. *Icarus* 45: 264–303.

Lucchitta BK (1982) Ice sculpture in the Martian outflow channels. *Journal of Geophysical Research* 87: 9951–9973.

Lucchitta BK (1984) Ice and debris in the fretted terrain, Mars. *Journal of Geophysical Research Supplement* 89: 409–418.

Lucchitta BK (1985) Geomorphologic evidence for ground ice on Mars. In: Klinger J, Benest D, Dollfus A, and Smoluchowski A (eds.) *Ices in the Solar System*, pp. 583–604. Dordrecht: D. Reidel Publishing.

Lucchitta BK (1987) Valles Marineris, Mars – Wet debris flows and ground ice. *Icarus* 72: 411–429.

Lucchitta BK (2001) Antarctic ice streams and outflow channels on Mars. *Geophysical Research Letters* 28: 403–406. http://dx.doi.org/10.1029/2000GL011924.

Lucchitta BK, Ferguson HM, and Summers C (1986) Sedimentary deposits in the northern lowland plains, Mars. *Journal of Geophysical Research* 91: 166–174.

Lucey P, Korotev RL, Gillis JJ, et al. (2006) New views of the moon. In: Jolliff BL, Wieczorek MA, Shearer CK, and Neal CR (eds.) *Understanding the Lunar Surface and Space–Moon Interactions*. *Reviews in Mineralogy and Geochemistry*, vol. 60, pp. 83–219. Corpus Christi, TX: World Scientific.

Luhmann JG (1997) Correction to 'The ancient oxygen exosphere of Mars implications for atmospheric evolution'. *Journal of Geophysical Research* 102: 1637.

Luhmann JG, Johnson RE, and Zhang MHG (1992) Evolutionary impact of sputtering of the Martian atmosphere by O^+ pick-up ions. *Geophysical Research Letters* 19: 2151–2154.

Luhmann JG and Kozyra JU (1991) Dayside pickup oxygen ion precipitation at Venus and Mars: Spatial distributions, energy deposition and consequences. *Journal of Geophysical Research* 96: 5457–5467.

Luo W and Stepinski TF (2006) Topographically derived maps of valley networks and drainage density in the Mare Tyrrhenum quadrangle of Mars. *Geophysical Research Letters* 33: L18202. http://dx.doi.org/10.1029/2006GL027346.

Mackay JR (1973) The growth of pingos, western Arctic coast, Canada. *Canadian Journal of Earth Sciences* 10(6): 979–1004.

Mackay JR (1979) Pingos of the Tuktoyaktuk Peninsula area, Northwest Territories. *Géographie Physique et Quaternaire* 33: 3–61.

Malin MC and Carr MH (1999) Groundwater formation of Martian valleys. *Nature* 397: 589–591.

Malin MC and Edgett KS (1999) Oceans or seas in the Martian northern lowlands: High-resolution imaging tests of proposed coastlines. *Geophysical Research Letters* 26: 3049–3052.

Malin MC and Edgett KS (2000a) Evidence for recent groundwater seepage and surface runoff on Mars. *Science* 288: 2330–2335.

Malin MC and Edgett KS (2001) Mars Global Surveyor Mars Orbiter camera: Interplanetary cruise through primary mission. *Journal of Geophysical Research* 106: 23429–23570.

Malin MC and Edgett KS (2003) Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* 302: 1931–1934.

Malin MC, Edgett KS, Posiolova LV, McColley SM, and Noe Dobrea EZ (2006) Present-day impact cratering rate and contemporary gully activity on Mars. *Science* 314: 1573–1577.

Manga M (2004) Martian floods at Cerberus Fossae can be produced by groundwater discharge. *Geophysical Research Letters* 31: 2702. http://dx.doi.org/10.1029/2003GL018958.

Mangold N (2003) Geomorphic analysis of lobate debris aprons on Mars at Mars Orbiter camera scale: Evidence for ice sublimation initiated by fractures. *Journal of Geophysical Research* 108(E4): 8021. http://dx.doi.org/10.1029/2002JE001885.

Mangold N (2005) High latitude patterned grounds on Mars: Classification, distribution and climatic control. *Icarus* 174: 336–359.

Mangold N and Allemand P (2001) Topographic analysis of features related to ice on Mars. *Geophysical Research Letters* 28: 407–410. http://dx.doi.org/10.1029/2000GL008491.

Mangold N, Costard F, and Forget F (2003) Debris flows over sand dunes on Mars: Evidence for liquid water. *Journal of Geophysical Research* 108: 5027. http://dx.doi.org/10.1029/2002JE001958.

Mangold N, Kite ES, Kleinhans MG, et al. (2012) The origin and timing of fluvial activity at Eberswalde crater, Mars. *Icarus* 220: 530–551. http://dx.doi.org/10.1016/j.icarus.2012.05.026.

Mangold N, Mangeney A, Migeon V, et al. (2010) Sinuous gullies on Mars: Frequency, distribution, and implications for flow properties. *Journal of Geophysical Research* 115: E11001. http://dx.doi.org/10.1029/2009JE003540.

Mangold N, Maurice S, Feldman WC, Costard F, and Forget F (2004a) Spatial relationships between patterned ground and ground ice detected by the neutron spectrometer on Mars. *Journal of Geophysical Research* 109(E8): E08001. http://dx.doi.org/10.1029/2004JE002235.

Mangold N, Quantin C, Ansan V, Delacourt C, and Allemand P (2004b) Evidence for precipitation on Mars from dendritic valleys in the Valles Marineris Area. *Science* 305: 78–81.

Manning R (1891) On the flow of water in open channels and pipes. *Transactions of the Institute of Civil Engineers, Ireland* 20: 161–207.

Margot JL, Campbell DB, Jurgens RH, and Slade MA (1999) Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* 284: 1658–1660.

Mars Channel Working Group (1983) Channels and valleys on Mars. *Geological Society of America Bulletin* 94: 1035–1054.

Masursky H (1973) An overview of geological results from *Mariner 9*. *Journal of Geophysical Research* 78: 4009–4030.

Masursky H, Boyce JM, Dial AL, Schaber GG, and Strobell ME (1977) Classification and time of formation of Martian channels based on Viking data. *Journal of Geophysical Research* 82: 4016–4038.

Masursky H, Chapman, MG, Dial Jr. AL, and Strobell ME (1986) Episodic channeling by volcanic flows in Mangala Valles region, Mars. *Reports of Planetary Geology Program, NASA TM-88383*, pp. 459–461.

Max MD and Clifford SM (2001) Initiation of Martian outflow channels: Related to the dissociation of gas hydrate? *Geophysical Research Letters* 28: 1787–1790.

McCauley JF (1978) *US Geological Survey Miscellaneous Investigation Series*. *Map I-897: Geological map of the coprates quadrangle of Mars*. Flagstaff, AZ: US Geological Survey.

McCord TB and Adams JB (1969) Spectral reflectivity of Mars. *Science* 163: 1058–1060.

McEwen AS, Eliason EM, and Bergstrom JW (1989) Mobility of large rock avalanches: Evidence from Valles Marineris. *Geology* 17: 1111–1114.

McEwen AS, Eliason EM, Bergstrom JW, et al. (2007a) Mars reconnaissance orbiter's high resolution imaging science experiment (HiRISE). *Journal of Geophysical Research* 112: E05S02. http://dx.doi.org/10.1029/2005JE002605.

McEwen AS, Hansen CJ, Delamere WA, et al. (2007b) A closer look at water-related geologic activity on Mars. *Science* 317: 1706–1709.

McEwen AS, Ojha L, Dundas CM, et al. (2011) Seasonal flows on warm Martian slopes. *Science* 333: 740–743.

McGill GE (1985) Age and origin of large Martian polygons. In: *16th Lunar Planetary Science Conference*, pp. 534–535.

McGill GE (1986) The giant polygons of Utopia, northern Martian plains. *Geophysical Research Letters* 13: 705–708.

McSween HY Jr. (1994) What we have learned about Mars from SNC meteorites. *Meteoritics* 29: 757–779.

Mellon MT and Jakosky BM (1993) Geographic variations in the thermal and diffusive stability of ground ice on Mars. *Journal of Geophysical Research* 98: 3345–3364.

Mellon MT and Jakosky BM (1995) The distribution and behavior of Martian ground ice during past and present epochs. *Journal of Geophysical Research* 100: 11781–11800.

Mellon MT and Phillips RJ (2001) Recent gullies on Mars and the source of liquid water. *Journal of Geophysical Research* 106: 23165–23180.

Menzies J (ed.) (1995) *Modern Glacial Environments – Processes, Dynamic and Sediments*, p. 621. Oxford: Butterworth-Heinemann.

Meslin PY, Gasnault O, Forni O, et al. (2013) Soil diversity and hydration as observed by ChemCam at Gale Crater, Mars. *Science* 341: 6153. http://dx.doi.org/10.1126/science.1238670.

Meunier A (2005) *Clays*. Berlin, Germany: Springer, 472 pp.

Mierdel K, Keppler H, Smyth JR, and Langenhorst F (2007) Water solubility in aluminous orthopyroxene and the origin of Earth's asthenosphere. *Science* 315: 364–368.

Milkovich SM, Head JW, and Marchant DR (2006) Debris-covered piedmont glaciers along the northwest flank of the Olympus Mons scarp: Evidence for low-latitude ice accumulation during the Late Amazonian of Mars. *Icarus* 181: 388–407.

Milliken RE, Mustard JF, and Goldsby DL (2003) Viscous flow features on the surface of Mars: Observations from high-resolution Mars Orbiter camera (MOC) images. *Journal of Geophysical Research* 108: 5057. http://dx.doi.org/10.1029/2002JE002005.

Milliken RE, Swayze GA, Arvidson RE, et al. (2008) Opaline silica in young deposits on Mars. *Geology* 36: 847–850.

Milton DJ (1974) Carbon dioxide hydrate and floods on Mars. *Science* 183: 654–655.

Mischna MA, Richardson MI, Wilson RJ, and McCleese DJ (2003) On the orbital forcing of Martian water and CO_2 cycles: A general circulation model study with simplified volatile schemes. *Journal of Geophysical Research* 108: 5062. http://dx.doi.org/10.1029/2003JE002051.

Mitrofanov I, Anfimov D, Kozyrev A, et al. (2002) Maps of subsurface hydrogen from the high energy neutron detector, Mars Odyssey. *Science* 297: 78–81.

Mitrofanov I, Sanin AB, Boynton WV, et al. (2010) Hydrogen mapping of the lunar south pole using the LRO neutron detector experiment LEND. *Science* 330: 483. http://dx.doi.org/10.1126/science.1185696.

Mitrofanov IG, Zuber MT, Litvak ML, et al. (2003) CO_2 snow depth and subsurface water-ice abundance in the northern hemisphere of Mars. *Science* 300: 2081–2084.

Miyamoto H, Dohm JM, Baker VR, Beyer RA, and Bourke M (2004) Dynamics of unusual debris flows on Martian sand dunes. *Geophysical Research Letters* 31: L13701. http://dx.doi.org/10.1029/2004GL020313.

Moore JM and Howard AD (2005) Large alluvial fans on Mars. *Journal of Geophysical Research* 110: E04005. http://dx.doi.org/10.1029/2004JE002352.

Moore JM, Howard AD, Dietrich WE, and Schenk PM (2003) Layered fluvial deposits: Implications for Noachian climate scenarios. *Geophysical Research Letters* 30: 2292. http://dx.doi.org/10.1029/2003GL019002.

Morgan GA, Head JW, Forget F, Madeleine J-B, and Spiga A (2010a) Gully formation on Mars: Two recent phases of formation suggested by links between morphology, slope orientation and insolation history. *Icarus* 208: 658–666.

Morgan GA, Head JW, and Marchant DR (2010b) Preservation of Late Amazonian Mars ice and water-related deposits in a unique crater environment in Noachis Terra: Age relationships between lobate debris tongues and gullies. *Icarus* 211: 347–365.

Morgan GA, Head JW, and Marchant DR (2011) Preservation of Late Amazonian Mars ice and water-related deposits in a unique crater environment in Noachis Terra: Age relationships between lobate debris tongues and gullies. *Icarus* 211: 347–365. http://dx.doi.org/10.1016/j.icarus.2010.08.004.

Morgenstern A, Hauber E, Reiss D, et al. (2007) Deposition and degradation of a volatile-rich layer in Utopia Planitia and implications for climate history on Mars. *Journal of Geophysical Research* 112: E06010. http://dx.doi.org/10.1029/2006JE002869.

Morris RV, et al. (2010) Identification of carbonate-rich outcrops on Mars by the Spirit rover. *Science* 329: 421–424.

Moses JI, Rawlins K, Zahnle K, and Dones L (1999) External sources of water for Mercury's putative ice deposits. *Icarus* 137: 197–221.

Mouginis-Mark P (1979) Martian fluidized crater morphology – Variations with crater size, latitude, altitude, and target material. *Journal of Geophysical Research* 84: 8011–8022.

Mouginis-Mark PJ, Wilson L, Head JW, Brown SH, Hall JL, and Sullivan KD (1984) Elysium Planitia, Mars – Regional geology, volcanology, and evidence for volcano–ground ice interactions. *Earth, Moon, and Planets* 30: 149–173.

Mouginot J, Pommerol A, Beck P, Kofman W, and Clifford SM (2012) Dielectric map of the Martian northern hemisphere and the nature of plain filling materials. *Geophysical Research Letters* 39: L02202. http://dx.doi.org/10.1029/2011GL050286.

Murase T (1962) Viscosity and related properties of volcanic rocks at 800° to 1400°C. *Journal of Faculty of Science* 487–584, Hokkaido University; Series 7, Geophysics.

Murchie SL, et al. (2009) The compact reconnaissance imaging spectrometer for Mars investigation and data set from the Mars Reconnaissance Orbiter's primary science phase. *Journal of Geophysical Research* 114: E00D07. http://dx.doi.org/10.1029/2009JE003344.

Murchie S, Mustard J, Bishop J, Head J, Pieters C, and Erard S (1993) Spatial variations in the spectral properties of bright regions on Mars. *Icarus* 105: 454–468.

Murphy DM and Koop T (2005) Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarternary Journal of the Royal Metereological Society* 608: 1539–1565.

Murray JB, Muller J, Neukum G, et al. (2005) Evidence from the Mars Express high resolution stereo camera for a frozen sea close to Mars' equator. *Nature* 434: 352–356.

Musselwhite DS, Swindle TD, and Lunine JI (2001) Liquid CO_2 breakout and the formation of recent small gullies on Mars. *Geophysical Research Letters* 28: 1283–1286.

Mustard JF, Cooper CD, and Rifkin MK (2001) Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* 412: 411–414.

Mustard JF, et al. (2008) Hydrated silicate minerals on Mars observed by the Mars reconnaissance orbiter CRISM instrument. *Nature* 454: 305–309.

Nachon M, Mangold N, Clegg S, et al. (2013) Sulfate calcium veins observed by the ChemCam instrument onboard Curiosity. *EPSC Abstracts* 8, EPSC2013-534.

Neukum G and Hiller K (1981) Martian ages. *Journal of Geophysical Research* 86: 3097–3121.

Neumann G, Cavanaugh JF, Sun X, et al. (2013) Bright and dark polar deposits on Mercury: Evidence for surface volatiles. *Science* 339(2013): 296–300. http://dx.doi.org/10.1126/science.1229764.

Newkirk G Jr. (1981) Solar variability on time scales of 10^5 years to $10^{9.6}$ years. *Geochimica et Cosmochimica Acta Supplement* 13: 293–301.

Newsom H (2005) Clays in the history of Mars. *Nature* 438: 570–571.

Newsom HE, Brittelle GE, Hibbitts CA, Crossey LJ, and Kudo AM (1996) Impact crater lakes on Mars. *Journal of Geophysical Research* 101: 14951–14956.

Nozette S, Lichtenberg CL, Spudis P, et al. (1996) The Clementine bistatic radar experiment. *Science* 274: 1495–1498.

Nozette S, Spudis PD, Robinson MS, Bussey DBJ, Lichtenberg C, and Bonner R (2001) Integration of lunar polar remote-sensing data sets: Evidence for ice at the lunar south pole. *Journal of Geophysical Research* 106: 23253–23266.

Nummedal D (1978) The role of liquefaction in channel development on Mars. *Reports of Planetary Geology Program, NASA TM-79729*, pp. 257–259.

Nummedal D and Prior DB (1981) Generation of Martian chaos and channels by debris flows. *Icarus* 45: 77–86.

Olhoeft GR (1981) Electrical properties of rocks. In: Touloukian YS, Judd WR, and Roy RF (eds.) *Physical Properties of Rocks and Minerals*, pp. 257–330. New York: McGraw-Hill.

Ori GG, Marinangeli L, and Baliva A (2000) Terraces in Gilbert-type deltas in crater lakes in Ismenius Lacus and Memnonia (Mars). *Journal of Geophysical Research* 105: 17629–17643.

Ori GG and Mosangini C (1998) Complex depositional systems in Hydrates Chaos, Mars: An example of sedimentary process interactions in the Martian hydrological cycle. *Journal of Geophysical Research* 103: 22713–22724.

Osterloo MM, Anderson FS, Hamilton VE, et al. (2010) Geologic context of proposed chloride-bearing materials on Mars. *Journal of Geophysical Research* 115: E10012. http://dx.doi.org/10.1029/2010JE003613.

Osterloo MM, Hamilton VE, Bandfield JL, et al. (2008) Chloride-bearing materials in the southern highlands of Mars. *Science* 319: 1651. http://dx.doi.org/10.1126/science.1150690.

Page DP and Murray JB (2006) Stratigraphical and morphological evidence for pingo genesis in the Cerberus plains. *Icarus* 183: 46–54.

Paige DA, Herkenhoff KE, and Murray BC (1990) Mariner 9 observations of the south polar CAP of Mars – Evidence for residual CO_2 frost. *Journal of Geophysical Research* 95: 1319–1335.

Paige D, Siegler M, and Harmon J (2013) Thermal stability of volatiles in the north polar region of Mercury. *Science* 141: 179–193.

Paige DA, Wood SE, and Vasavada AR (1992) The thermal stability of water ice at the poles of Mercury. *Science* 258: 643–646.

Parker TJ (1989) Channels and valley networks associated with Argyre Planitia, Mars. In: *Lunar and Planetary Institute Conference*, pp. 826–827.

Parker TJ, Gorsline DS, Saunders RS, Pieri DC, and Schneeberger DM (1993) Coastal geomorphology of the Martian northern plains. *Journal of Geophysical Research* 98: 11061–11078.

Parker TJ, Saunders RS, and Schneeberger DM (1989) Transitional morphology in west Deuteronilus Mensae, Mars – Implications for modification of the lowland/upland boundary. *Icarus* 82: 111–145.

Paterson WSB (1994) *The Physics of Glaciers*, 3rd edn. Oxford: Pergamon, 480 pp.

Pechmann JC (1980) The origin of polygonal troughs on the northern plains of Mars. *Icarus* 42: 185–210. http://dx.doi.org/10.1016/0019-1035(80)90071-8.

Pelletier JD, Kolb KJ, McEwen AS, and Kirk RL (2008) Recent bright gully deposits on Mars: Wet or dry flow? *Geology* 36: 211–214.

Penido JC, Fassett CI, and Som SM (2013) Scaling relationships and concavity of small valley networks on Mars. *Planetary and Space Science* 75: 105–116. http://dx.doi.org/10.1016/j.pss.2012.09.009.

Perron JT, Dietrich WE, Howard AD, McKean JA, and Pettinga JR (2003) Ice-driven creep on Martian debris slopes. *Geophysical Research Letters* 30: 1747. http://dx.doi.org/10.1029/2003GL017603.

Phillips RJ, Zuber MT, Smrekar SE, et al. (2008) Mars north polar deposits: Stratigraphy, age, and geodynamical response. *Science* 320: 1182–1185.

Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.

Picardi G, Plaut JJ, Biccari D, et al. (2005) Radar soundings of the subsurface of Mars. *Science* 310: 1925–1928.

Pieri D (1976) Distribution of small channels on the Martian Surface. *Icarus* 27: 25–50.

Pieters CM, Goswami JN, Clark RN, et al. (2009) Character and spatial distribution of OH/H_2O on the surface of the Moon seen by M3 on Chandrayaan-1. *Science* 326. http://dx.doi.org/10.1126/science.1178658.

Pollack JB (1991) Kuiper prize lecture: Present and past climates of the terrestrial planets. *Icarus* 91: 173–198.

Pollack JB, Kasting JF, Richardson SM, and Poliakoff K (1987) The case for a wet, warm climate on early Mars. *Icarus* 71: 203–224.
Pollack JB, Roush T, Witteborn F, et al. (1990) Thermal emission spectra of Mars (5.4–10.5 microns) – Evidence for sulfates, carbonates, and hydrates. *Journal of Geophysical Research* 95: 14595–14627.
Pondrelli M, Baliva A, Di Lorenzo S, Marinangeli L, and Rossi AP (2005) Complex evolution of paleolacustrine systems on Mars: An example from the Holden crater. *Journal of Geophysical Research* 110: E04016. http://dx.doi.org/10.1029/2004JE002335.
Pondrelli M, Rossi AP, Marinangeli L, et al. (2008) Evolution and depositional environments of the Eberswalde fan delta, Mars. *Icarus* 197: 429–451. http://dx.doi.org/10.1016/j.icarus.2008.05.018.
Porsild AE (1938) Earth mounds in unglaciated Arctic northwest America. *Geographical Review* 28: 46–58.
Poulet F, Bibring J-P, Mustard JF, et al. (2005) Phyllosilicates on Mars and implications for early Martian climate. *Nature* 438: 623–627.
Raack J, Reiss D, and Hiesinger H (2012) Gullies and their relationships to the dust-ice mantle in the northwestern Argyre Basin, Mars. *Icarus* 219: 129–141.
Rabette M, McKay C, Pilewskie P, and Young R (2002) The runaway greenhouse effect on Earth. In: *Astrobiology Science Conference NASA Ames Research Center, Moffett Field, CA, 8 April* .
Reedy RC (1978) Planetary gamma-ray spectroscopy. In: *Lunar and Planetary Science Conference,* pp. 2961–2984.
Reiss D, Erkeling G, Bauch KE, and Hiesinger H (2010) Evidence for present day gully activity on the Russell crater dune field, Mars. *Geophysical Research Letters* 37: L06203. http://dx.doi.org/10.1029/2009GL042192.
Reiss D, Hauber E, Hiesinger H, et al. (2011) Terrestrial gullies and debris flows tracks on Svalbard as planetary analogs for Mars. *Geological Society of America, Special Paper* 483: 165–175.
Reiss D, Hauber E, Michael G, Jaumann R, and Neukum G (2005) Small rampart craters in an equatorial region on Mars: Implications for near-surface water or ice. *Geophysical Research Letters* 32: L10202. http://dx.doi.org/10.1029/2005GL022758.
Reiss D, Hiesinger H, Hauber E, and Gwinner K (2009) Regional differences in gully occurrence on Mars: A comparison between the Hale and Bond craters. *Planetary and Space Science* 57: 958–974.
Reiss D and Jaumann R (2003) Recent debris flows on Mars: Seasonal observations of the Russell crater dune field. *Geophysical Research Letters* 30(6): 1321. http://dx.doi.org/10.1029/2002GL016704.
Reiss D, Jaumann R, Kereszturi A, Sik A, and Neukum G (2007) Gullies and avalanche scars on Martian dark dunes. In: *Proceedings of the 38th Lunar and Planetary Science Conference, 12–16 March, Houston, Texas, abstract 1993*.
Reiss D, van Gasselt S, Hauber E, Michael G, Jaumann R, and Neukum G (2006) Ages of rampart craters in equatorial regions on Mars: Implications for the past and present distribution of ground ice. *Meteoritics and Planetary Science* 41: 1437–1452.
Reiss D, van Gasselt S, Neukum G, and Jaumann R (2004) Absolute dune ages and implications for the time of formation of gullies in Nirgal Vallis, Mars. *Journal of Geophysical Research* 109: E06007. http://dx.doi.org/10.1029/2004JE002251.
Richardson MI and Mischna MA (2005) Long-term evolution of transient liquid water on Mars. *Journal of Geophysical Research* 110: E03003. http://dx.doi.org/10.1029/2004JE002367.
Rieder R, Economou T, Wänke H, et al. (1997) The chemical composition of Martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* 278: 1771–1774.
Rieder R, Gellert R, Brückner J, et al. (2003) The new Athena alpha particle X-ray spectrometer for the Mars exploration rovers. *Journal of Geophysical Research* 108: 8066. http://dx.doi.org/10.1029/2003JE002150.
Rivkin AS, Howell EG, Vilas F, and Lebofsky LA (2002) Hydrated minerals on asteroids: The astronomical record. *NASA STI/Recon Technical Report N, 2, 46,798*.
Robinson MS and Tanaka KL (1990) Magnitude of a catastrophic flood event at Kasei Vallis, Mars. *Geology* 18: 902–905.
Rodriguez JAP, Sasaki S, and Miyamoto H (2003) Nature and hydrological relevance of the Shalbatana complex underground cavernous system. *Geophysical Research Letters* 30: 1304. http://dx.doi.org/10.1029/2002GL016547.
Rodriguez JAP, Tanaka KL, Miyamoto H, and Sasaki S (2006) Nature and characteristics of the flows that carved the Simud and Tiu outflow channels, Mars. *Geophysical Research Letters* 33: L08S04. http://dx.doi.org/10.1029/2005GL024320.
Rossbacher LA and Judson S (1981) Ground ice on Mars – Inventory, distribution, and resulting landforms. *Icarus* 45: 39–59.
Roush TL, Blaney DL, and Singer RB (1993) The surface composition of Mars as inferred from spectroscopic observations. In: Pieters CM and Englert PAJ (eds.) *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pp. 367–393. Cambridge: Cambridge University Press.
Ruesch O, Poulet F, Vincendon M, et al. (2012) Compositional investigation of the proposed chloride-bearing materials on Mars using near-infrared orbital data from OMEGA/MEx. *Journal of Geophysical Research* 117: E00J13. http://dx.doi.org/10.1029/2012JE004108.
Ruff SW (2004) Spectral evidence for zeolite in the dust on Mars. *Icarus* 168: 131–143.
Russell PS and Head JW (2003) Elysium-Utopia flows as mega-lahars: A model of dike intrusion, cryosphere cracking, and water-sediment release. *Journal of Geophysical Research* 108(E6): 5064. http://dx.doi.org/10.1029/2002JE001995.
Ryan AJ and Christensen PR (2012) Coils and polygonal crust in the Athabasca Valles region, Mars, as evidence for a volcanic history. *Science* 336: 449–452. http://dx.doi.org/10.1126/science.1219437.
Salisbury JW, D'Aria DM, and Jarosewich E (1991) Midinfrared (2.5–13.5 microns) reflectance spectra of powdered stony meteorites. *Icarus* 92: 280–297.
Schmincke H-U (2000) *Vulkanismus*. Darmstadt: Wissenschaftliche Buchgesellschaft, 264 pp.
Schon SC, Head JW, and Fassett CI (2009) Unique chronostratigraphic marker in depositional fan stratigraphy on Mars: Evidence for ca. 1.25 Ma gully activity and surficial meltwater origin. *Geology* 37: 207–210.
Schorghofer N and Aharonson O (2005) Stability and exchange of subsurface ice on Mars. *Journal of Geophysical Research* 110: E05003. http://dx.doi.org/10.1029/2004JE002350.
Schultz PH (1992) Atmospheric effects on ejecta emplacement. *Journal of Geophysical Research* 97: 11623–11662.
Schultz PH and Gault DE (1979) Atmospheric effects on Martian ejecta emplacement. *Journal of Geophysical Research* 84: 7669–7687.
Scott DH and Dohm JM (1992) Mars highland channels: An age reassessment. *Lunar and Planetary Science* 23: 1251–1252.
Scott DH, Dohm JM, and Rice JW (1995) *Geological Survey Miscellaneous Investigation Series Map, I-2461-A: Map Showing Channels and Possible Paleolake Basins*. Flagstaff, AZ: US Geological Survey.
Scott DH and Tanaka KL (1986) *US Geological Survey Miscellaneous Investigation Series Map, I-1802-A: Geologic map of the western equatorial region of Mars*. Flagstaff, AZ: US Geological Survey.
Sears DWG and Chittenden JD (2005) On laboratory simulation and the temperature dependence of the evaporation rate of brine on Mars. *Geophysical Research Letters* 32: L23203. http://dx.doi.org/10.1029/2005GL024154.
Seibert NM and Kargel JS (2001) Small-scale Martian polygonal terrain: Implications for liquid surface water. *Geophysical Research Letters* 28: 899–902.
Séjourné A, Costard F, Gargani J, Soare RJ, Fedorov A, and Marmo C (2011) Scalloped depressions and small-sized polygons in western Utopia Planitia, Mars: A new formation hypothesis. *Planetary and Space Science* 59: 412–422.
Shean DE, Head JW, and Marchant DR (2005) Origin and evolution of a cold-based tropical mountain glacier on Mars: The Pavonis Mons fan-shaped deposit. *Journal of Geophysical Research* 110: 5001. http://dx.doi.org/10.1029/2004JE002360.
Shinbrot T, Duong NH, Kwan L, and Alvarez MM (2004) Dry granular flows can generate surface features resembling those seen in Martian gullies. *Proceedings of the National Academy of Sciences* 101: 8542–8546.
Shivola A (1996) *Electromagnetic Mixing Formulas and Applications*. London: IEE, 284 pp.
Simpson RA and Tyler GL (1999) Reanalysis of Clementine bistatic radar data from the lunar South Pole. *Journal of Geophysical Research* 104: 3845–3862.
Skok JR, Mustard JF, Ehlmann BL, et al. (2010) Silica deposits in the Nili Patera caldera on the Syrtis Major volcanic complex on Mars. *Nature Geoscience* 3: 838–841. http://dx.doi.org/10.1038/ngeo990.
Slade MA, Butler BJ, and Muhleman DO (1992) Mercury radar imaging – Evidence for polar ice. *Science* 258: 635–640.
Smith PH, Tamppari LK, Arvidson RE, et al. (2009) H_2O at the Phoenix landing site. *Science* 325: 58–61.
Smith DE, Zuber MT, Frey HV, et al. (2001) Mars Orbiter laser altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research* 106: 23689–23722.
Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1502.
Soare RJ, Bun Tseung JMW, and Peloquin C (2005) Possible thermokarst and alas formation in Utopia Planitia, Mars. *Lunar and Planetary Science* XXXVI: 1103 (CD-ROM).
Soderblom LA (1992) The composition and mineralogy of the Martian surface from spectroscopic observations – 0.3 micron to 50 microns. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 557–593. Tucson, AZ: University of Arizona Press.
Soderblom LA, Kreidler TJ, and Masursky H (1973) Latitudinal distribution of a debris mantle on the Martian surface. *Journal of Geophysical Research* 78: 4117–4122.

Solomon SC, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.

Soukhovitskaya V and Manga M (2006) Martian landslides in Valles Marineris: Wet or dry? *Icarus* 180: 348–352. http://dx.doi.org/10.1016/j.icarus.2005.09.008.

Sparks RSJ, Gardeweg MC, Calder ES, and Matthews SJ (1997) Erosion by pyroclastic flows on Lascar Volcano, Chile. *Bulletin of Volcanology* 58: 557–565.

Squyres SW (1978) Martian fretted terrain – Flow of erosional debris. *Icarus* 34: 600–613.

Squyres SW (1979) The distribution of lobate debris aprons and similar flows on Mars. *Journal of Geophysical Research* 84: 8087–8096.

Squyres SW (1989) Urey prize lecture – Water on Mars. *Icarus* 79: 229–288. http://dx.doi.org/10.1016/0019-1035(89)90078-X.

Squyres SW, Arvidson RE, Baumgartner ET, et al. (2003) Athena Mars rover science investigation. *Journal of Geophysical Research* 108: 8062. http://dx.doi.org/10.1029/2003JE002121.

Squyres SW and Carr MH (1986) Geomorphic evidence for the distribution of ground ice on Mars. *Science* 231: 249–252.

Squyres SW, Clifford SM, Kuzmin RO, Zimbelman JR, and Costard FM (1992) Ice in the Martian regolith. In: Kieffer HH, Jakosky BM, Synder CM, and Matthews MS (eds.) *Mars*, pp. 523–554. Tucson, AZ: University of Arizona Press.

Squyres SW, Grotzinger JP, Arvidson RE, et al. (2004) *In situ* evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* 306: 1709–1714.

Squyres SW and Kasting JF (1994) Early Mars – How warm and how wet. *Science* 265: 744–749.

Squyres SW, Knoll AH, Arvidson RE, et al. (2006) Two years at Meridiani Planum: Results from the Opportunity rover. *Science* 313: 1403–1407.

Stacey NJS, Campbell DB, and Ford PG (1997) Arecibo radar mapping of the lunar poles: A search for ice deposits. *Science* 276: 1527–1530.

Stepinski TF and Coradetti S (2004) Comparing morphologies of drainage basins on Mars and Earth using integral-geometry and neural maps. *Geophysical Research Letters* 31: L15604. http://dx.doi.org/10.1029/2004GL020359.

Stepinski TF and Stepinski AP (2005) Morphology of drainage basins as an indicator of climate on early Mars. *Journal of Geophysical Research* 110: E12S12. http://dx.doi.org/10.1029/2005JE002448.

Stewart ST and Nimmo F (2002) Surface runoff features on Mars: Testing the carbon dioxide formation hypothesis. *Journal of Geophysical Research* 107: 5069. http://dx.doi.org/10.1029/2000JE001465.

Stewart ST, O'Keefe JD, and Ahrens TJ (2001) The relationship between Rampart crater morphologies and the amount of subsurface ice. *Lunar and Planetary Science* XXXII: 2092 (CD-ROM).

Stockstill KR, Moersch JE, Ruff SW, Baldridge A, and Farmer J (2005) Thermal emission spectrometer hyperspectral analyses of proposed paleolake basins on Mars: No evidence for in-place carbonates. *Journal of Geophysical Research* 110: 10004. http://dx.doi.org/10.1029/2004JE002353.

Sugden DE and John BS (1976) *Glaciers and Landscape*. London: Arnold, 328 pp.

Summerfield MA (1991) *Global Geomorphology*. London: Longman, 535 pp.

Sunshine JM, Farnham TL, Feaga LM, et al. (2009) Temporal and spatial variability of lunar hydration as observed by the Deep Impact spacecraft. *Science* 326. http://dx.doi.org/10.1126/science.1179788.

Tanaka KL (1997) Sedimentary history and mass flow structures of Chryse and Acidalia Planitiae, Mars. *Journal of Geophysical Research* 102: 4131–4150.

Tanaka KL (1999) Debris-flow origin for the Simud/Tiu deposit on Mars. *Journal of Geophysical Research* 104: 8637–8652.

Tanaka KL, Banerdt WB, Kargel JS, and Hoffman N (2001) Huge, CO_2-charged debris-flow deposit and tectonic sagging in the northern plains of Mars. *Geology* 29: 427–430.

Terada N, Machida S, and Shinagawa H (2002) Global hybrid simulation of the Kelvin–Helmholtz instability at the Venus ionopause. *Journal of Geophysical Research* 107(A12): 1471–1490.

Theilig E and Greeley R (1979) Plains and channels in the Lunae Planum-Chryse Planitia region of Mars. *Journal of Geophysical Research* 84: 7994–8010.

Thollot P, Mangold N, Ansan V, et al. (2012) Most Mars minerals in a nutshell: Various alteration phases formed in a single environment in Noctis Labyrinthus. *Journal of Geophysical Research* 117: E00J06.

Thomas P, Squyres S, Herkenhoff K, Howard A, and Murray B (1992) *Polar Deposits of Mars, in Mars*. Tucson, AZ: University of Arizona Press, pp. 767–795.

Tokano T and Bish DL (2005) Hydration state and abundance of zeolites on Mars and the water cycle. *Journal of Geophysical Research* 110: E12S08. http://dx.doi.org/10.1029/2005JE002410.

Toon OB, Pollack JB, Ward W, Burns JA, and Bilski K (1980) The astronomical theory of climatic change on Mars. *Icarus* 44: 552–607.

Tosca NJ and Knoll AH (2009) Juvenile chemical sediments and the long-term persistence of water at the surface of Mars. *Earth and Planetary Science Letters* 286: 379–386. http://dx.doi.org/10.1016/j.epsl.2009.07.004.

Toulmin P, Rose HJ, Christian RP, et al. (1977) Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *Journal of Geophysical Research* 82: 4625–4634.

Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.

Treiman AH (2003) Geologic settings of Martian gullies: Implications for their origins. *Journal of Geophysical Research* 108: 8031. http://dx.doi.org/10.1029/2002JE001900.

Twidale CR (1976) *Analysis of Landforms*. Sydney: Wiley, 572 pp.

Ulrich M, Morgenstern A, Günther F, et al. (2010) Thermokarst in Siberian ice-rich permafrost: Comparison to asymmetric scalloped depressions on Mars. *Journal of Geophysical Research* 115: E10009. http://dx.doi.org/10.1029/2010JE003640.

Urey HC (1967) Water on the Moon. *Nature* 216: 1094.

Vallance JW (2000) Lahars. In: Sigurdsson H (ed.) *Encyclopedia of Volcanoes*, pp. 601–616. San Diego, CA: Academic Press.

van Everdingen RO (2005) *Mult-Language Glossary of Permafrost and Related Ground-Ice Terms*. Calgary: International Permafrost Association, http://nsidc.org/fgdc/glossary (accessed on February 2007).

van Gasselt S, Hauber E, and Neukum G (2007) Cold-climate modification of Martian landscapes: A case study of a spatulate debris landform in the Hellas Montes Region, Mars. *Journal of Geophysical Research* 112: 2156–2202. http://dx.doi.org/10.1029/2006JE002842.

van Gasselt S, Reiss D, Thorpe AK, and Neukum G (2005) Seasonal variations of polygonal thermal contraction crack patterns in a south polar trough, Mars. *Journal of Geophysical Research, Planets* 110: 8002. http://dx.doi.org/10.1029/2004JE002385.

Vaniman DT, Bish DL, Ming DW, et al. (2014) Mineralogy of a mudstone at Yellowknife Bay, Gale Crater, Mars. *Science* 343: 1243480. http://dx.doi.org/10.1126/science.1243480.

Vaniman DT, Chipera SJ, Bish DL, Carey JW, Fialips CI, and Feldman WC (2004) Sulfate salts, regolith interactions, and water storage in equatorial Martian regolith. *Lunar and Planetary Science* XXXV: 1690 (CD-ROM).

Vasavada AR, Paige DA, and Wood SE (1999) Near-surface temperatures on Mercury and the moon and the stability of polar ice deposits. *Icarus* 141: 179–193.

Védie E, Costard F, Font M, and Lagarde JL (2008) Laboratory simulations of Martian gullies on sand dunes. *Geophysical Research Letters* 35: L21501. http://dx.doi.org/10.1029/2008GL035638.

Vialov SS (1958) Regularities of glacial shields movements and the theory of plastic viscous flow. *Physics of the Movements of Ice, International Association of Hydraulic Sciences* 47: 266–275.

Wahrhaftig C and Cox A (1959) Rock glaciers in the Alaska Range. *Geological Society of America Bulletin* 70: 383–436.

Wallace P and Anderson A (2000) Volatiles in magmas. In: Sigurdsson H (ed.) *Encyclopedia of Volcanoes*, pp. 149–170. New York: Academic Press.

Wang A, Korotev RL, Jolliff BL, et al. (2006) Evidence of phyllosilicates in Wooly Patch, an altered rock encountered at West Spur, Columbia Hills, by the Spirit rover in Gusev crater, Mars. *Journal of Geophysical Research* 111: E02S16. http://dx.doi.org/10.1029/2005JE002516.

Wänke H, Brückner J, Dreibus G, Rieder R, and Ryabchikov I (2001) Chemical composition of rocks and soils at the pathfinder site. *Space Science Reviews* 96: 317–330.

Ward WR (1973) Large-scale variations in the obliquity of Mars. *Science* 181: 260–262.

Ward WR (1974) Climatic variations on Mars. I: Astronomical theory of insolation. *Journal of Geophysical Research* 79: 3375–3386.

Warner N, Gupta S, Lin S-Y, et al. (2010) Late Noachian to Hesperian climate change on Mars: Evidence of episodic warming from transient crater lakes near Ares Vallis. *Journal of Geophysical Research* 115: E06013. http://dx.doi.org/10.1029/2009JE003522.

Washburn AL (1979) *Geocryology*, 2nd edn. London: Arnold, 406 pp.

Watson K, Murray BC, and Brown H (1961) The behavior of volatiles on the lunar surface. *Journal of Geophysical Research* 66: 3033–3045.

Webb VE (2004) Putative shorelines in northern Arabia Terra, Mars. *Journal of Geophysical Research* 109: 9010. http://dx.doi.org/10.1029/2003JE002205.

Weitz CM, Irwin RP, Chuang FC, Bourke MC, and Crown DA (2006) Formation of a terraced fan deposit in Coprates Catena, Mars. *Icarus* 184: 436–451.

Werner SC (2006) *Major Aspects of the Chronostratigraphy and Geologic Evolutionary History of Mars*. Dissertation, Freie Universität Berlin, 160S.

Werner SC, van Gasselt S, and Neukum G (2003) Continual geological activity in Athabasca Valles, Mars. *Journal of Geophysical Research* 108: 8081. http://dx.doi.org/10.1029/2002JE002020.

Whalley WB and Azizi F (2003) Rock glaciers and protalus landforms: Analogous forms and ice sources on Earth and Mars. *Journal of Geophysical Research* 108: 8032. http://dx.doi.org/10.1029/2002JE001864.

Wharton RA, Crosby JM, McKay CP, and Rice JW (1995) Paleolakes on Mars. *Journal of Paleolimnology* 13: 267–283.

Williams RS (1978) Geomorphoc processes in Iceland and on Mars: A comparative appraisal from orbital images. *Geological Society of America Abstract Programs* 10: 517.

Williams RME, Grotzinger JP, Dietrich WE, et al. (2013) Martian fluvial conglomerates at Gale Crater. *Science* 340: 1068–1072. http://dx.doi.org/10.1126/science.1237317.

Williams H and McBirney AR (1979) *Volcanology*. San Francisco, CA: Freeman & Cooper, 397 pp.

Williams RM, Phillips RJ, and Malin MC (2000) Flow rates and duration within Kasei Valles, Mars: Implications for the formation of a Martian ocean. *Geophysical Research Letters* 27: 1073–1076.

Williams KE, Toon OB, and Heldmann J (2007) Modeling water ice lifetimes at recent Martian gully locations. *Geophysical Research Letters* 34.

Williams KE, Toon OB, Heldmann JL, and Mellon MT (2009) Ancient melting of mid-latitude snowpacks on Mars as a water source for gullies. *Icarus* 200: 418–425.

Wilson L, Ghatan GJ, Head JW, and Mitchell KL (2004) Mars outflow channels: A reappraisal of the estimation of water flow velocities from water depths, regional slopes, and channel floor properties. *Journal of Geophysical Research* 109: 9003. http://dx.doi.org/10.1029/2004JE002281.

Wood BE, Muller HR, Zank G, and Linsky JL (2002) Measured mass loss rates of solar like stars as a function of age and activity. *Astrophysical Journal* 574: 412–425.

Woods AW (1995) The dynamics of explosive volcanic eruptions. *Reviews of Geophysics* 33: 495–530.

Woronow A (1981) Preflow stresses in Martian rampart ejecta blankets – A means of estimating the water content. *Icarus* 45: 320–330.

Wyatt MB and McSween HY (2002) Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars. *Nature* 417: 263–266.

Yershov ED (1998) *General Geocryology*. Cambridge: Cambridge University Press, 580 pp.

Yoshikawa K (2003) Origin of the polygons and the thickness of Vastitas Borealis Formation in Western Utopia Planitia on Mars. *Geophysical Research Letters* 30(12): 1603. http://dx.doi.org/10.1029/2003GL017165.

Yung YL and Pinto JP (1978) Primitive atmosphere and implications for the formation of channels on Mars. *Nature* 288: 735–738.

Zahnle KJ and Walker JCG (1982) The evolution of solar ultraviolet luminosity. *Reviews of Geophysics and Space Physics* 20: 280–292.

Zanetti M, Hiesinger H, Reiss D, Hauber E, and Neukum G (2010) Distribution and evolution of scalloped terrain in the southern hemisphere, Mars. *Icarus* 206: 691–706.

Zent AP (1999) An open, snow-based, hydrologic system as an analog for Noachian Mars. *Lunar and Planetary Science* XXX: 1803 (CD-ROM).

Zent AP, Howard DJ, and Quinn RC (2001) H_2O adsorption on smectites: Application to the diurnal variation of H_2O in the Martian atmosphere. *Journal of Geophysical Research* 106: 14667–14674.

Zent AP and Quinn RC (1995) Simultaneous adsorption of CO_2 and H_2O under Mars-like conditions and application to the evolution of the Martian climate. *Journal of Geophysical Research* 100: 5341–5349.

Zent AP and Quinn RC (1997) Measurement of H_2O adsorption under Mars-like conditions: Effects of adsorbent heterogeneity. *Journal of Geophysical Research* 102: 9085–9096.

Zhang MHG, Luhmann JG, Bougher SW, and Nagy AF (1993) The ancient oxygen exosphere of Mars implication for atmospheric evaluations. *Journal of Geophysical Research* 98: 10915–10923.

Zimbelman JR (2001) Image resolution and evaluation of genetic hypotheses for planetary landscapes. *Geomorphology* 37: 179–199.

This page intentionally left blank

10.12 Solid Planet–Atmosphere Interactions

M Yu Zolotov, Arizona State University, Tempe, AZ, USA

© 2015 Elsevier B.V. All rights reserved.

10.12.1	**Atmosphere–Surface Interplay on Solar System Bodies**	411
10.12.2	**Observational Constraints on Venus and Mars**	411
10.12.2.1	Physical and Chemical Signs of Surface Rock Alteration	412
10.12.2.2	Atmospheric Compositions in the Context of Atmosphere–Surface Interactions	414
10.12.3	**Chemistry of Atmosphere–Surface Reactions on Mars and Venus**	415
10.12.3.1	Redox Processes	415
10.12.3.1.1	Oxidation on Venus	415
10.12.3.1.2	Oxidation on Mars	416
10.12.3.2	Reactions with Sulfur Gases	417
10.12.3.3	Reactions with H_2O	419
10.12.3.4	Reactions with CO_2	419
10.12.3.5	Reactions with Halogen-Bearing Species	420
10.12.3.6	Condensation of Metal and Nonmetal Species: The Role of Altitude	421
10.12.3.7	Kinetic Aspects of Surface–Gas-Type Reactions	421
10.12.3.8	Local Reactions Caused by Impacts and Volcanic Eruptions	421
10.12.4	**Wind-Related Processes**	422
10.12.5	**Atmosphere–Surface Interactions Throughout History**	422
10.12.6	**Summary and Unsolved Questions**	423
References		424

10.12.1 Atmosphere–Surface Interplay on Solar System Bodies

Even without an abundant aqueous phase, planetary surfaces are coupled with planetary atmospheres through complex physical and chemical processes. Volcanic degassing, impacts, and seepage from mantle and crustal rocks supply gases into atmospheres. The condensation of released volatiles affects abundances of atmospheric gases and surface composition. Compositions of atmospheres and surfaces are influenced by chemical interactions of gases with surface rocks and liquids, if they are present. In some cases, atmospheric abundances of volatiles can be controlled (buffered) by condensation/sublimation equilibria or chemical reactions at the surface. The rates of chemical reactions at the atmosphere–surface interface are affected by temperatures that, in turn, are influenced by solar luminosity, cloudiness, and the greenhouse effect. Eolian processes contribute to weathering, mass transfer, and separation of surface solids. Throughout history, atmosphere–surface interactions could have been affected by changing solar luminosity, orbital parameters, internal processes, impact cratering, geologic resurfacing, and gas escape into space.

Three major terrestrial planets – Venus, Earth, and Mars – reveal a broad-scale coupling of atmospheric and solid-body processes. On Earth, the major pathways of endogenic and exogenic processes are mainly determined by mantle convection, plate tectonics and corresponding volcanism, the hydrosphere, life, and impacts. On Mars and Venus, surface composition and climatic evolution are affected by volcanism, impacts, and atmosphere–surface interactions with limited and/or temporal influences from aqueous processes. This chapter is mainly devoted to atmosphere–surface interactions on Venus and Mars.

Despite significant differences in surface temperature, pressure, and atmospheric composition (Table 1), the atmosphere–surface interactions on Venus and Mars have several common characteristics. Aqueous solutions are unstable on both of their surfaces. On both planets, surfaces have not been significantly affected by volcanic, impact, and eolian processes for hundreds of millions of years. Nevertheless, surface materials reveal signs of alteration by coupled physical and chemical processes. Chemical alteration could have included oxidation of primary ferrous (Fe^{2+}-bearing) minerals and sulfides and consumption of sulfur, halogens, and water from the atmosphere. Physical processes include the condensation of species with different volatilities, wind abrasion, generation of fine-grained surface materials, mechanical separation of mineral grains, and formation of eolian features, including layered deposits with low porosity. Atmosphere–surface interactions are altitude- and latitude-dependent, owing to the variations in surface temperature, near-surface pressure, and concentrations of atmospheric gases. On both planets, the composition of atmosphere and surface materials reveals past geologic events, including intensive volcanic/impact degassing and resurfacing, as well as processes that occurred under wetter conditions.

10.12.2 Observational Constraints on Venus and Mars

This section presents data on surface materials and atmospheres in the context of gas–solid-type interactions. More information on planetary surfaces and atmospheres can be found in *Treatise on Geochemistry*, (e.g., Fegley, 2014; McSween and McLennan, 2014).

Table 1 Environmental conditions and the composition of major and chemically active gases in lower atmospheres of Venus and Mars

	Venus[a]	*Mars*[b]
Temperature	740 K	210 K
Pressure	95.6 bar	5.6×10^{-3} bar
CO_2	$96.5 \pm 0.8\%$	95.3%
N_2	3.5 ± 0.8	$2.7 \pm 0.5\%$
^{40}Ar	31^{+20}_{-10} ppm	$1.6 \pm 0.3\%$
$^{36+38}Ar$	35.5^{+20}_{-10} ppm	5.3 ± 1 ppm
O_2	–[c]	$0.13 \pm 0.01\%$
SO_2	150 ± 30 ppm (22–42 km)	<0.1 ppb
H_2O	30 ± 15 ppm (5–45 km)	0.03%[d]
CO	17 ± 1 ppm (17 km)	0.08%
COS	4.4 ± 1 ppm (33 km)	<0.01 ppm
H_2S	3 ± 2 ppm (<20 km)	<0.1 ppm
HCl	0.4 ± 0.03 ppm (0–74 km)	<2 ppb
S_{1-8}	20 ppb (<50 km)	–
HF	5 ± 3 ppb (35–70 km)	–
H_2	–	15 ± 5 ppm
CH_4	–	<1.3 ppb
H_2O_2	–	0–40 ppb 18 ± 0.4 ppb
O_3	–	0.04–0.2 ppm
NO	5.5 ± 1.5 ppb (<60 km)	<1.7 ppb
D/H	0.016 ± 0.002[e] 0.019 ± 0.006 (<50 km)	$(7.8 \pm 0.3) \times 10^{-4}$

ppm, parts per million by volume; ppb, parts per billion by volume.
[a]Venus temperature and pressure correspond to the modal planetary radius (6051.4 km). Gas abundances represent values closer to the surface. Concentrations of CO and COS are latitude- and altitude-dependent. Near-surface concentrations of these and other gases may be different.
[b]Martian surface temperature and pressure are estimates of the global annual averages (Kieffer et al., 1992).
[c]Near the surface, O_2 pressure is evaluated as 10^{-21} bar (Fegley et al., 1997b; Zolotov, 1996).
[d]Variable with season and location.
[e]D/H ratios for Venus represent Pioneer Venus in situ measurements and ground spectroscopy, respectively.
Sources: Esposito et al. (1997), Encrenaz et al. (2012), Fegley (2014), Krasnopolsky and Feldman (2001), Krasnopolsky et al. (2004), Krasnopolsky (2010), Krasnopolsky (2011), McSween and McLennan (2014), Webster et al. (2013), and Owen (1992).

Table 2 Chemical composition of Venus' surface materials at the Venera and Vega landing sites (mass %)

Oxide	*Venera 13*	*Venera 14*	*Vega 2*
SiO_2	45.1 ± 3.0	48.7 ± 3.6	45.6 ± 3.2
Al_2O_3	15.8 ± 3.0	17.9 ± 2.6	16.0 ± 1.8
FeO	9.3 ± 2.2	8.8 ± 1.8	7.7 ± 1.1
MnO	0.2 ± 0.1	0.16 ± 0.08	0.14 ± 0.12
MgO	11.4 ± 6.2	8.1 ± 3.3	11.5 ± 3.7
CaO	7.1 ± 0.96	10.3 ± 1.2	7.5 ± 0.7
K_2O	4.0 ± 0.63	0.2 ± 0.07	0.1 ± 0.08
TiO_2	1.59 ± 0.45	1.25 ± 0.41	0.2 ± 0.1
SO_3	1.62 ± 1.0	0.88 ± 0.77	4.7 ± 1.5
Cl	<0.3	<0.4	<0.3
Total	96.1	96.3	93.4

The data were obtained with XRF analysis. Uncertainties are $\pm 1\delta$. All Fe is given as FeO. Na content has not been measured.
Sources: Surkov et al. (1984) and Surkov et al. (1986).

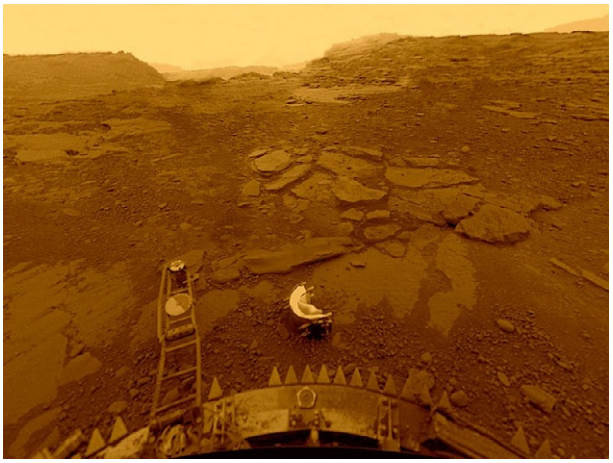


Figure 1 Layered rocks and rock fragments at the surface of Venus at the landing site of Venera 13. The surface is black at Venus' conditions and reddish color represents the surface at room temperature (cf. Pieters et al., 1986). The upper part of the image has been artistically created by Don Mitchell using other Venus' surface images.

10.12.2.1 Physical and Chemical Signs of Surface Rock Alteration

Basalts are the more abundant igneous rocks on Mars and Venus. On Mars, the mafic composition of primary rocks is inferred from studies of Martian meteorites; in situ analyses by the Viking, Mars Exploration Rover (MER), and Mars Science Laboratory (the Curiosity rover) missions; and remote sensing data (Banin et al., 1992; Blake et al., 2013; Clark et al., 1982; McSween and McLennan, 2014; McSween et al., 2006a,b, 2008; Mustard et al., 2005; Rieder et al., 2004). The mainly basaltic composition of Venus' plains is inferred from the morphology of widespread volcanic features observed in Venera 15–16 (Barsukov et al., 1986) and Magellan (Crumpler et al., 1997) radar images and x-ray fluorescence (XRF) analysis at the landing sites of Venera 13–14 and Vega 2 (**Table 2**). Gamma-ray spectrometric analysis of K, U, and Th at Venera 9–10 and Vega 1–2 landing sites is consistent with mafic rocks, while Venera 8 materials resemble alkaline rocks (Surkov et al., 1987).

Despite their chiefly basaltic composition, surface rocks are physically and chemically altered. Planetary surfaces are partially covered by wind-blown material, rock fragments, and fine-grained soil. Porous layered deposits have been observed at the landing sites on Venus (e.g., Basilevsky et al., 1985; Florensky et al., 1977, 1983a,b; Garvin et al., 1984), at Meridiani Planum on Mars (Squyres et al., 2006), and at the Columbia Hills in Gusev crater (**Figures 1** and **2**). It has been suggested that layered rocks at these sites were deposited from the atmosphere by eolian activity, explosive volcanism, and/or impact events (Basilevsky et al., 1985, 2004; Garvin et al., 1984; Grotzinger et al., 2005; Knauth et al., 2005).

Layered rocks observed at Venera 13–14 landing sites are mechanically weak materials with high porosities (50–60%), low densities (1.4–1.5 g cm^{-2}, Florensky et al., 1983a,b), and low bearing capacities (4–5 kg cm^{-2}, Avduevsky et al., 1983; 2.6–10 kg cm^{-2}, Kemurdzhian et al., 1983). These rocks resist drilling similar to weathered porous basalt or ashy tuff (Barmin and Shevchenko, 1983). Layered rocks observed at the

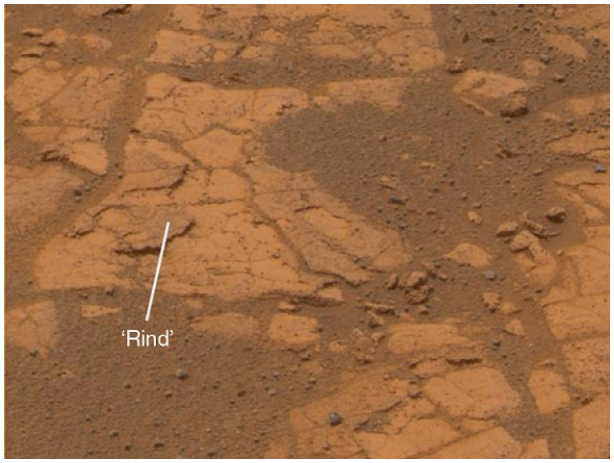


Figure 2 Layered sedimentary rocks at the Meridiani Planum, Mars. The rind-like features appear to cap the rocks and are probably more resistant to wind erosion than surrounding rocks. The image was taken with the Opportunity rover during 552 sol (13 August 2005). Photo credit NASA/JPL/Cornell.

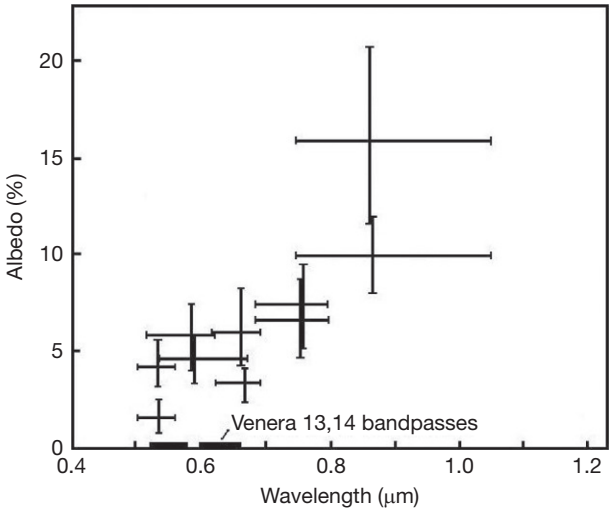


Figure 3 Reflectance properties of the Venus' surface at the Venera 9–10 landing sites (Ekonomov et al., 1980; Pieters et al., 1986). Horizontal bars indicate the width at half height of the Venera filters. Reproduced from Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383, with permission from AAAs.

Venera 9–10 landing sites have higher bearing capacities (30–300 kg cm^{-2}), although they possess properties analogous to volcanic tuffs and/or weathered basalts (Basilevsky et al., 2004; Kemurdzhian et al., 1983). Predominantly horizontal layering and some signs of cross-bedding (Venera 10) are consistent with deposition from the atmosphere followed by mild lithification and erosion. Fine-grained material observed in Venera 9–10 and 13 panoramas is the likely product of the degradation of local rocks (Basilevsky et al., 2004; Florensky et al., 1977; Garvin et al., 1984).

Radar observations of Venus' volcanic plains by the Pioneer Venus, Venera 15–16, and Magellan orbiters revealed a dielectric constant (δ) of ~5, typical for basalts (Ford and Pettengill, 1983; Pettengill et al., 1997). These data are consistent with the electrical resistance of rocks at the landing site of Vega 2 (10^6 Ωm), which is common for heated basalts (Kemurdzhian et al., 1983). Rocks at the landing sites of Venera 13–14 had significantly lower electrical resistances (89 and 73 Ωm, respectively; Kemurdzhian et al., 1983). These measurements suggest lateral heterogeneity of surface material. On the highlands, ~4.5 km above the mean planetary radius (6051.5 km), surface materials are characterized by elevated radar reflectivity that corresponds to a dielectric constant as high as 20–30 (Ford and Pettengill, 1983; Pettengill et al., 1997). These high δ values are inconsistent with unaltered igneous rocks.

Martian layered rocks at Meridiani Planum (**Figure 2**) and Gusev crater (e.g., Home Plate in the Columbia Hills) are characterized by high porosity, low density, and sometimes cross-bedded structures (e.g., Fergason et al., 2006; Grotzinger et al., 2005). The thermal inertia of soils and bedforms at both MER landing sites is consistent with particle diameters between 45 and 160 μm (silt to fine sand; Fergason et al., 2006), which agrees with the grain size of the most easily suspended particles (100–150 μm; Greeley et al., 1980). The finest grains could represent a globally homogenized material that has been observed at the Viking and Mars Pathfinder landing sites (Bell et al., 2000; Christensen and Moore, 1992). In addition, rocks, soils, and individual sand grains are at least partially coated by particles with a diameter of 0.2–10 μm, which represents atmospheric dust.

On Mars and Venus, surface materials exhibit specific absorptions in the visible and near-infrared spectral ranges, indicating the presence of Fe^{3+}-bearing (ferric) species, which are uncommon in unaltered basalts. Reflectance spectra of Venus' surface measured at the Venera 9–10 landing sites (Ekonomov et al., 1980) roughly match with the spectrum of the ferric oxide hematite (α-Fe_2O_3) heated to Venus' surface temperatures (Pieters et al., 1986), as can be seen in **Figures 3** and **4**. On Mars, the absorption of light by ferric oxides is responsible for the reddish hue of dust, soils, and rock coatings (Bell et al., 2000; Soderblom, 1992). This inference is consistent with the detection of goethite (α-FeOOH), hematite, nanophase ferric oxide/oxyhydroxide, ferric sulfates, and clays by thermal emission (Christensen et al., 2001; Glotch et al., 2006) and Mössbauer (Klingelhöfer et al., 2004; Morris et al., 2006, 2008) spectroscopies. Only minor alteration observed in Martian meteorites (Bridges et al., 2001; Gooding, 1992), which were excavated by impacts, implies that oxidation occurs mainly at the Martian surface.

Chemical analyses of surface materials on both planets reveal incorporation of volatile elements into alteration products. Venus' surface rocks (**Table 2**) are significantly enriched in S compared with basalts, which typically contain <0.2 wt% S. Martian soils (**Table 3**) are enriched in S, Cl, Br, and bound H_2O compared with normal basalts. Rock fragments possess enrichments in S and halogens, indicating dust coatings and/or weathering rinds (Gellert et al., 2006; Rieder et al., 1997, 2004). The variable abundances of volatiles imply their exogenic source. Elevated abundances of sulfur observed in Venus' and Mars' surface rocks may indicate global processes of basalt enrichment in near-surface environments.

10.12.2.2 Atmospheric Compositions in the Context of Atmosphere–Surface Interactions

Carbon dioxide and N_2 are the major atmospheric components on Mars and Venus (Table 1). CO_2 is a major component of terrestrial volcanic gases (Oppenheimer, 2003; Symonds et al., 1994) and could be an abundant product of mantle outgassing on other terrestrial planets. On Venus, the mass of atmospheric CO_2 is similar to the mass of CO_2 stored in terrestrial crustal carbonates (Table 4). Compared to Earth, this suggests that the majority of CO_2 degassed on Venus resides in its atmosphere. On Mars, the low mass of CO_2 in the atmosphere, even when normalized to planetary mass, may indicate early incorporation of CO_2 in carbonates and/or escape of CO_2 into space.

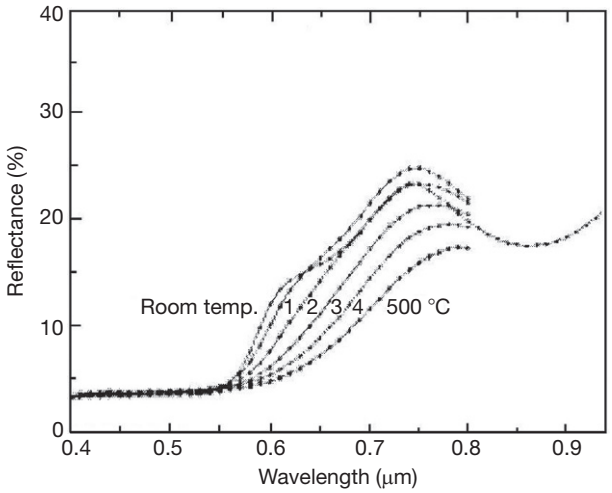


Figure 4 Laboratory reflectance spectra of the ferric oxide hematite at temperatures ranging from room temperature to 500 °C. Reproduced from Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383, with permission from AAAs.

On both planets, active trace species play a major role in atmospheric chemical processes and rock alteration. On Mars, photochemically generated species (e.g., CO, O_2, H_2, H_2O_2, O_3, HO_2, O, OH, H, NO, and NO_2) are distributed throughout the lower atmosphere and can affect the surface composition (Krasnopolsky, 2011; McElroy et al., 1977; Yung and DeMore, 1999). On Venus, highly oxidized photochemically derived species are subjected to thermochemical reactions below clouds (48–60 km) and may not reach the surface. Near the surface, atmospheric chemistry is controlled by thermochemical reactions, with an exception of photochemical interaction between S_3 and S_4 gases. In contrast, in the Martian atmosphere, thermochemical processes are insignificant because of low densities and temperatures.

On Mars, densities of atmospheric gases vary with season and latitude. Nonsymmetrical seasonal condensation of atmospheric CO_2 in polar regions causes variations in total atmospheric pressure (James et al., 1992). Similarly, the abundance and partial pressure (P) of water vapor are strongly dependent on season and latitude, reaching the highest values in spring in the Northern Hemisphere (Jakosky and Haberle, 1992; Smith, 2002).

On Venus, altitudinal variations in concentrations of trace gases are caused by temperature–pressure-dependent

Table 3 Chemical composition of Martian soils detected in situ (in mass %)

Oxide	*Viking 1, Chryse Planitia*	*Viking 2, Utopia Planitia*	*Average Viking soil*	*Average Mars Pathfinder soil*	*Average Opportunity, Meridiani Planum*	*Average Spirit, Gusev crater*	*Curiosity, Rocknest Portage, Gale crater*
SiO_2	43	43	43.4^{+6}_{-6}	48.6 ± 2.5	45.7 ± 1.3	46.1 ± 0.9	42.88 ± 0.47
Al_2O_3	7.3	7[a]	7.2^{+4}_{-4}	8.3 ± 0.8	9.25 ± 0.50	10.19 ± 0.69	9.43 ± 0.14
Fe_2O_3	18.5	17.8	18.2^{+5}_{-2}	17.5 ± 1.8	18.8 ± 1.2[b]	16.3 ± 1.1[b]	19.19 ± 0.12[b]
MnO	–	–	–	–	0.37 ± 0.02	0.32 ± 0.03	0.41 ± 0.01
MgO	6	6[a]	6.0^{+5}_{-3}	7.5 ± 1.2	7.38 ± 0.29	8.67 ± 0.60	8.69 ± 0.14
CaO	5.9	5.7	5.8^{+2}_{-2}	6.3 ± 1.0	6.93 ± 0.32	6.30 ± 0.29	7.28 ± 0.07
Na_2O	–	–	–	2.2 ± 0.9	2.21 ± 0.18	3.01 ± 0.30	2.72 ± 0.10
K_2O	<0.15	<0.15	<0.15	0.3 ± 0.1	0.48 ± 0.05	0.44 ± 0.07	0.49 ± 0.01
TiO_2	0.66	0.56	$0.6^{+0.25}_{-0.25}$	1.1 ± 0.2	1.03 ± 0.12	0.88 ± 0.19	1.19 ± 0.03
Cr_2O_3	–	–	–	–	0.41 ± 0.06	0.33 ± 0.07	0.49 ± 0.02
P_2O_5	–	–	–	–	0.84 ± 0.06	0.91 ± 0.31	0.94 ± 0.03
SO_3	6.6	8.1	7.4^{+6}_{-2}	5.4 ± 1.1	5.83 ± 1.04	5.78 ± 1.25	5.45 ± 0.10
Cl	0.7	0.5	$0.8^{+1.5}_{-0.5}$	0.6 ± 0.2	0.65 ± 0.09	0.70 ± 0.16	0.69 ± 0.02
Br, ppm[c]	–	–	–	–	100 ± 111	53 ± 46	26 ± 6
Ni, ppm	–	–	–	–	457 ± 97	476 ± 142	446 ± 29
Zn, ppm	–	–	–	–	309 ± 87	270 ± 90	337 ± 17
Total	88.7	88.7	89.4	97.8	99.88	99.88	99.85

[a]Not measured and assumed to be the same as in Viking 1 soils.
[b]For Opportunity and Spirit data, all Fe is reported as FeO.
[c]ppm, parts per million by mass.
Sources: Clark et al. (1982), Banin et al. (1992), Bell et al. (2000), Rieder et al. (1997), and Blake et al. (2013).

Table 4 Masses of volatiles in other shells of Venus, Mars, and Earth (g)

Gas	*Venus atmosphere*	*Mars atmosphere*	*Earth atmosphere, hydrosphere, and crust*[a]
CO_2	4.6×10^{23}	2.1×10^{19}	4.1×10^{23}
N_2	1.1×10^{22}	7.6×10^{17}	5.3×10^{21}
H_2O	7.0×10^{18}	6.5×10^{15}	2.3×10^{24}
S	5.4×10^{19}	$<2 \times 10^{12}$	2.6×10^{22}
Cl	1.6×10^{17}	$<2 \times 10^{12}$	3.6×10^{22}
Abundances normalized to planetary masses (g/g)			
CO_2	9.45×10^{-5}	3.3×10^{-8}	6.86×10^{-5}
N_2	2.3×10^{-6}	1.2×10^{-9}	8.9×10^{-7}
H_2O	1.4×10^{-9}	1.0×10^{-11}	3.9×10^{-4}
S	1.1×10^{-8}	$<3 \times 10^{-15}$	4.4×10^{-6}
Cl	3.3×10^{-11}	$<3 \times 10^{-15}$	6.6×10^{-6}

[a]Terrestrial data are from Ronov and Yaroshevsky (1976) and Volkov (1992).

thermochemical reactions and Hadley circulation beneath the cloud deck. Thermochemical processes below the clouds create concentration gradients for several species. For example, thermochemical interaction of CO with S-bearing gases to form COS could be implied by the observed decreasing CO and increasing COS abundances toward the surface (Krasnopolsky, 2007; Krasnopolsky and Pollack, 1994). The upward motion of gases in the equatorial region and downwellings in pole regions in the Hadley cell could produce higher concentrations of CO in high latitudes and elevated abundances of COS in equatorial regions.

At Venus' surface, catalysis by solid phases and sluggish air circulation in the boundary layer may allow thermochemical equilibration at least among some gases. For example, a model of chemical equilibration among CO_2, CO, COS, and SO_2 at the surface is consistent with observations (Fegley et al., 1997b; Krasnopolsky and Pollack, 1994; Zolotov, 1996). Concentrations of several other near-surface trace gases may be controlled by thermochemical equilibria as well (**Table 5**). The possibility of equilibration among gases increases below the surface, where migration of gases is controlled by diffusion through pore spaces. High porosity facilitates gas–surface interactions in the near-surface layer. Without new data from Venus, assuming thermochemical equilibrium provides a way to evaluate concentrations of undetected gases. Above Venus' surface, especially at elevations of several kilometers, equilibrium predictions are not consistent with observations (e.g., Krasnopolsky, 2007; Krasnopolsky and Pollack, 1994; Zolotov, 1996). Disequilibria among major chemically active gases (CO_2, SO_2, CO, and COS) in Venus' highlands constrain the evaluation of atmospheric composition and pathways of gas–mineral interactions.

10.12.3 Chemistry of Atmosphere–Surface Reactions on Mars and Venus

Atmospheric–surface reactions have been considered since the early 1960s in attempts to interpret spacecraft and telescopic observations, as reviewed by Von Zahn et al. (1983), Lewis and Prinn (1984), Volkov et al. (1986), Banin et al. (1992), Fegley and Treiman (1992), Gooding et al. (1992), Zolotov and Volkov (1992), Fegley et al. (1997a), Wood (1997), and Fegley (2014). Despite many theoretical and some experimental efforts, there is no consensus among planetologists on the stability of minerals, as well as rates and timing of processes.

Table 5 Composition of the near-surface atmosphere of Venus according to a thermochemical equilibrium model for conditions at the modal planetary radius (6051.4 km, 740 K, and 95.6 bar), in volume fraction

Gas	*Mole ratio*	*Gas*	*Mole ratio*
CO_2	0.965	H_2S	1.1×10^{-7}
N_2	0.035	H_2	3.5×10^{-9}
SO_2	1.3×10^{-4}	S_3	8.5×10^{-10}
H_2O	3.0×10^{-5}	CS_2	6.5×10^{-11}
COS	2.8×10^{-5}	SO	3.1×10^{-11}
CO	1.7×10^{-5}	S_4	7.0×10^{-12}
S_2	2.3×10^{-7}	O_2	$10^{-21.3}$

Source: Zolotov MYu (1996) A model for the thermal equilibrium of the surface venusian atmosphere. *Geochemistry International* 33(10): 80–100.

10.12.3.1 Redox Processes

Chemical reactions involving electron transfers (redox processes) are the major class of reactions at the atmosphere–surface interface on Venus and Mars. Photochemical production of oxidized atmospheric gases (e.g., SO_3, O_2, and O_3) and volcanic supply of CO_2, SO_2, and H_2O create the possibility for oxidation of freshly exposed igneous rocks. The observed oxidizing nature of surface materials on Mars and Venus (see **Section 10.12.2.1**) implies that atmospheric oxidants were present and involved. On Mars and Venus, surface rocks are only partially oxidized. On both planets, a major phase of mineral oxidation could have occurred during earlier epochs that were warmer and more humid (see **Section 10.12.5**).

10.12.3.1.1 Oxidation on Venus

The atmospheric redox state near the mean planetary radius on Venus corresponds to conditions at which the iron oxides magnetite ($Fe^{2+}Fe^{3+}{}_2O_4$ and Fe_3O_4) and hematite coexist with each other (Fegley et al., 1997b; Zolotov, 1996), consistent with the interpretation of near-infrared reflectance patterns of rocks observed at the Venera 9–10 landing sites (Pieters et al., 1986). An oxidized, hematite-bearing surface

environment agrees with the results from the color experiments on the Venera 13–14 landers that provided a lower limit for atmospheric CO of 2.3–6.8 ppm (Fegley et al., 1997b; Zolotov, 1996).

The strong absorption of light in the visible spectral range observed for Venera 9–10 and Venera 13–14 rocks (Figure 3) is consistent with the low albedo of magnetite and hematite at the surface temperature. The lack of Fe enrichment in analyzed rocks (Table 2) compared with typical basalts implies only a low-albedo coating of ferric oxides. An analysis of Venera 13 color data reveals an enrichment of surface fines in 'red' Fe^{3+}-bearing material (Shkuratov et al., 1987). This interpretation agrees with diffusion-limited oxidation by atmospheric gases by which smaller grains alter faster. In addition, glassy impact- or volcanic-generated components in Venus' fines should experience higher rates of alteration compared to crystalline minerals.

Magnetite can form as an intermediate phase during oxidation of ferrous silicates to hematite. Theoretical studies of mineral stability show that primary ferrous minerals in basalts (olivines and pyroxenes) could be oxidized by O-bearing atmospheric gases, leading to the production of reduced gases. The pathways of oxidation to magnetite can be exemplified by the following reactions:

$$3FeO + CO_2 \rightarrow Fe_3O_4 + CO \quad [1]$$

$$3FeO + 0.5SO_2 \rightarrow Fe_3O_4 + 0.25S_2 \quad [2]$$

$$3FeO + H_2O \rightarrow Fe_3O_4 + H_2 \quad [3]$$

Here, ferrous iron oxide is a component of rock-forming silicates such as olivine and pyroxene and silicate glasses. In addition to ferrous silicates, the primary iron sulfides troilite (FeS) and pyrrhotite ($Fe_{1-x}S$) can be oxidized to ferric oxides and/or pyrite (FeS_2), as can be seen in Figure 5.

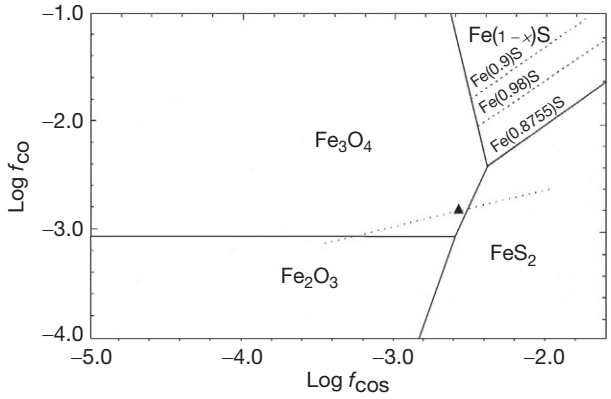


Figure 5 Stability of minerals in the Fe–S–O system at the conditions of Venus' surface (740 K and 95.6 bar total pressure). Activities of minerals are unity. The dotted line represents equilibrium conditions among CO_2, CO, COS, and SO_2 at CO_2 volume fraction of 0.965 and mole fraction of SO_2 of 130 ppm. The filled triangle represents equilibrium conditions at which mole fraction of CO is 17 ppm and volume fraction of COS is 28 ppm. The plot demonstrates that abundances of gases in the near-surface atmosphere (Table 5) are close to the conditions of magnetite–pyrite, hematite–magnetite, and hematite–magnetite–pyrite equilibria.

Pyrrhotite-to-pyrite conversions can occur through the uptake of H_2S and/or S_2 via reactions

$$Fe_{1-x}S + H_2S \rightarrow (1-x)FeS_2 + xS_2 + H_2 \quad [4]$$

$$Fe_{1-x}S + 0.5(1-2x)S_2 \rightarrow (1-x)FeS_2 \quad [5]$$

The correspondence of atmospheric redox conditions to the conditions of the magnetite–hematite equilibrium is not likely to be accidental and could reflect ongoing magnetite-to-hematite transition via some of the following reactions:

$$2Fe_3O_4 + CO_2 \rightarrow 3Fe_2O_3 + CO \quad [6]$$

$$2Fe_3O_4 + 0.5SO_2 \rightarrow 3Fe_2O_3 + 0.25S_2 \quad [7]$$

$$2Fe_3O_4 + H_2O \rightarrow 3Fe_2O_3 + H_2 \quad [8]$$

It is not clear which atmospheric gas plays the major role in rock oxidation. If equilibrium models for atmospheric chemistry are valid, all considered oxidation reactions have the same thermodynamic potential. One cannot tell which reaction is the major one without the consideration of reaction kinetics. The efficiency of basalt oxidation by CO_2 at Venus' surface conditions has been demonstrated experimentally (Fegley et al., 1995). Oxidation by SO_2 can occur through disproportionation of sulfur and release of S_2 (Terradellas and Bonnetain, 1973). Water is an oxidant of ferrous silicates under hot O_2-free conditions. More experiments are needed to evaluate the relative role of the considered oxidation reactions [1]–[8] in the preceding text. At different altitudes, temperature and compositional changes should affect the thermodynamics and kinetics of redox reactions. New observations and experiments are needed to understand redox processes at high altitudes.

The coexistence of magnetite and hematite at the surface may buffer atmospheric CO_2/CO, SO_2/COS, SO_2/S_2, and H_2O/H_2 ratios, which designate the oxidation state of the lower Venus' atmosphere. Buffering by reactions [6]–[8] means that the gas fugacity ratios remain unchanged until magnetite is consumed. Here, the magnetite-to-hematite conversion can occur only through oxidation by one or two gases, likely CO_2 and/or H_2O, followed by thermochemical re-equilibration among other gases. The dominance of ferrous minerals in mafic crustal rocks compared to the amount of oxidized gases and a slow rate of H_2 escape (Donahue et al., 1997) could have prevented profound oxidation of rocks to magnetite and hematite. In other words, slowly reacting Fe oxides may maintain the redox conditions (fO_2) in the near-surface atmosphere, where the majority of air mass is located.

10.12.3.1.2 Oxidation on Mars

The reddish color of the Martian surface and the detection of hematite, goethite, and nanophase Fe^{3+} oxide/oxyhydroxide; Ca-, Mg-, and Fe^{3+}-sulfates; and perchlorates at the surface of Mars (Hecht et al., 2009; Klingelhöfer et al., 2004; Morris et al., 2006, 2008; Leshin et al., 2013) and Martian orbital spectrometers (e.g., Christensen et al., 2001; Gendrin et al., 2005; Murchie et al., 2009) are consistent with near-surface oxidation.

As on Venus, Martian surface conditions thermodynamically favor oxidation of Fe^{2+} silicates and sulfides to Fe^{3+}-species and sulfates (Gooding, 1978; Gooding et al., 1992; O'Connor, 1968). On Mars, oxidation of minerals may be

driven by photochemically produced O_2, O_3, and H_2O_2 (e.g., Encrenaz et al., 2012; McElroy et al., 1977; Parkinson and Hunten, 1972). In addition, H_2O_2 can form by electrochemical processes in Martian dust devils and storms (Atreya et al., 2006; Delory et al., 2006). In all cases, peroxides are produced via dissociation of atmospheric O-bearing gases. The formation of perchlorates (Hecht et al., 2009; Leshin et al., 2013) could involve photochemistry-driven oxidation of Cl species in the atmosphere and/or at the surface (Catling et al., 2010). The presence of H_2O_2, perchlorates, and possibly other photochemically derived oxidants in the soil is consistent with the results of Viking biological experiments (Klein, 1978; Oyama and Berdahl, 1977, 1979; Zent and McKay, 1994) and initial results obtained by the Curiosity rover.

The mechanisms and environments of Martian surface oxidation could have varied in time and space and remain poorly known. Huguenin (1973a,b) reported experimental results showing rapid photooxidation of magnetite to hematite in an O_2-bearing atmosphere. Subsequent experiments by Clark et al. (1979) and Morris and Lauer (1980) failed to observe alteration of magnetite, as well as photostimulated alteration of other minerals. This inconsistency questioned results from theoretical works in which the effects of solar irradiation on the formation of goethite, clays, and other hydrated minerals were investigated (Huguenin, 1974, 1976). Similarly, the formation of peroxides in Martian soil through frost weathering of ferrous silicates (Huguenin, 1982; Huguenin et al., 1979) remains to be verified. Some insights could be obtained from the studies of rocks exposed in polar regions on Earth. In contrast to inorganic species, photochemical oxidation of organic compounds occurs readily under Martian conditions (e.g., Moores and Schuerger, 2012; Poch et al., 2013).

The occurrence of hematite-rich spherules at Meridiani Planum (**Figure 2**; Klingelhöfer et al., 2004), layered sulfate deposits at Valles Marineris and other locations (Gendrin et al., 2005; Langevin et al., 2005; Murchie et al., 2009; Roach et al., 2010), and localized goethite and sulfate enrichments in Gusev crater (Ming et al., 2006; Morris et al., 2008) reveals aqueous precipitation. It is possible that atmospheric and soil oxidants dissolved in surface aqueous solutions caused oxidation of reduced solutes (e.g., Fe^{2+}) and sulfide minerals via overall reactions

$$Fe^{2+} + 0.25O_2 + 1.5H_2O \rightarrow FeOOH + 2H^+ \quad [9]$$

$$FeS + 2.25O_2 + 1.5H_2O \rightarrow FeOOH + 2H^+ + SO_4^{2-} \quad [10]$$

where FeS represents iron sulfides and FeOOH stands for ferric oxides and oxyhydroxides. The presence of dissolved perchlorates and H_2O_2 increases the rate of oxidation. The contribution of ClO_4^- and/or H_2O_2 to aqueous oxidation might have been essential at low water/rock ratios that may represent episodic formation of water films or precipitation of atmospheric aerosols.

Reactions [9] and [10] produce protons that decrease the solution pH and the oxidation rate. In terrestrial environments, aqueous oxidation of sulfides leads to pH values as low as −2 (Nordstrom et al., 2000). On Mars, mineral precipitation in acidic near-surface aqueous environments is consistent with MER observations at Meridiani Planum and Gusev crater (Clark et al., 2005; Haskin et al., 2005; Ming et al., 2006; Morris et al., 2008). In particular, the detection of jarosite (a complex Fe^{3+} sulfate), in layered rocks at Meridiani, indicates low-pH, oxidizing conditions at the time of mineral formation (Klingelhöfer et al., 2004). The presence of jarosite shows that limited amounts of solution did not neutralize before evaporating and/or freezing. In other locations, surface aqueous processes could have occurred at higher water/rock ratios and led to complete neutralization, consistent with the detection of Fe–Mg phyllosilicates in Noachian deposits (Ehlmann et al., 2011; Murchie et al., 2009; Poulet et al., 2005). The composition of these phyllosilicates indicates their formation in reduced enthronements. The lack of atmospheric oxidants could have caused subsurface solutions to be reduced and alkaline in composition.

The scale of aqueous oxidation on Mars could have been limited by the mass and production rate of atmospheric oxidants. Periods of intensive volcanism may have been characterized by low concentrations of oxidants owing to their interaction with reduced volcanic gases. Although strong impacts could have generated an array of strong oxidants (see **Section 10.12.3.8**), the limited oxidation of surface materials is consistent with the widespread occurrence of olivine and Fe^{2+}-bearing pyroxenes in soils and exposed igneous rocks (e.g., Hoefen et al., 2003; Klingelhöfer et al., 2004; McSween et al., 2008, 2010; Mustard et al., 2005; Vaniman et al., 2014). Spectral and chemical observations of rock fragments and red-colored magnetite-rich dust collected on MER magnets (Goetz et al., 2005) imply that the formation of ferric oxides affected only the surfaces of mineral grains and rock fragments.

During periods characterized by a reducing but warm atmosphere, surface oxidation could have occurred by water via reactions [3] and [8]. The escape of H_2 produced through dissociation of H_2O would have contributed to net oxidation. Deep below the Martian surface, oxidation by liquid water may have occurred over Mars' history. Slow interaction of H_2 with dissolved carbonate species (HCO_3^- and CO_3^{2-}) can account for the formation of simple organic species.

10.12.3.2 Reactions with Sulfur Gases

Sulfur-bearing species are abundant in the Martian soil, rock coatings, and sedimentary deposits, as well as in Venus' rocks in which S was analyzed (**Tables 2** and **3**). In all cases, the S content does not correlate with rock composition, consistent with a secondary origin of sulfur. Despite the elevated S content, only a fraction of metals (Ca, Mg, and Fe) are bonded to sulfur, indicating incomplete alteration of surface rocks with respect to sulfur.

Although the mineralogy of S-bearing phases remains to be determined for many of Mars' and all of Venus' rocks, sulfates are the most likely secondary minerals. On Mars, the presence of sulfates is inferred from correlations with Mg (e.g., Bell et al., 2000; Clark et al., 1982, 2005), from infrared and Mössbauer observations by MER (e.g., Glotch et al., 2006; Klingelhöfer et al., 2004; Morris et al., 2006, 2008; Squyres et al., 2012), and from near-infrared spectra from the Martian orbiters (e.g., Gendrin et al., 2005; Langevin et al., 2005; Murchie et al., 2009; Roach et al., 2010). These data are consistent with the presence of Mg sulfates with various degrees of hydration (e.g., kieserite, $MgSO_4 \cdot H_2O$, and epsomite, $MgSO_4 \cdot 7H_2O$), gypsum

($CaSO_4 \cdot 2H_2O$), and ferric sulfates. Gypsum is present in some Martian meteorites (Bridges et al., 2001). Several other sulfates may be present as well.

On Venus, thermodynamic instability of rock-forming calcium silicates (plagioclases and pyroxenes) with respect to atmospheric SO_2 (**Table 1**) should drive the formation of anhydrite ($CaSO_4$) (Barsukov et al., 1982; Fegley and Treiman, 1992; Volkov et al., 1986). The formation of sulfates in O_2-free conditions can occur through disproportionation of SO_2 via gas–solid-type reactions

$$Ca_2Si_2O_6(\text{in pyroxene}) + 3SO_2 \rightarrow 2CaSO_4 + 2SiO_2 + 0.5S_2 \quad [11]$$

$$CaAl_2Si_2O_8(\text{in plagioclase}) + 1.5SO_2 \rightarrow CaSO_4 + 2SiO_2 + Al_2O_3 + 0.25S_2 \quad [12]$$

Alternatively or additionally, oxidation of S^{4+} (in SO_2) to S^{6+} (in sulfate) by atmospheric CO_2 (e.g., Fegley and Treiman, 1992) or H_2O may form $CaSO_4$, releasing CO or H_2, respectively. Although the presence of anhydrite in Venus' rocks has not been confirmed, conversion of diopside ($CaMgSi_2O_6$) to $CaSO_4$ has been observed in experiments at a temperature of 1106 K (Fegley and Prinn, 1989).

Despite a strong thermodynamic drive for anhydrite formation, the composition of Venus' rocks (**Table 2**) indicates incomplete sulfurization of Ca silicates (**Figure 6**). It can be seen that Venera 14 rocks are less altered with respect to S consumption. This is consistent with reduced physical weathering relative to Venera 13 rocks (Basilevsky et al., 1985; Garvin et al., 1984), which suggests younger rocks at the Venera 14 landing site. However, the timescale of SO_2 removal from the atmosphere is not known. It is not clear if current or recent volcanism is needed to replenish atmospheric SO_2.

In addition to sulfates, surface materials on both planets may contain secondary sulfides formed at the surface. On Mars, cooling of H_2S-rich hydrothermal fluids may cause precipitation of pyrite, as happens on Earth in oceanic 'black smokers.' Rare occurrences of liquid water at the surface and limited oxidants could have restricted subsequent oxidation of pyrite to ferric oxides and sulfates. In other words, both primary and secondary sulfides may be present in Martian surface materials, consistent with recent Curiosity data (McAdam et al., 2014; Vaniman et al., 2014). Likewise, thermodynamically unstable olivine, Fe–Mg pyroxenes, and magnetite are present in Martian soils and dust (Goetz et al., 2005; Hoefen et al., 2003; Klingelhöfer et al., 2004; McSween et al., 2010; Morris et al., 2006, 2008; Mustard et al., 2005).

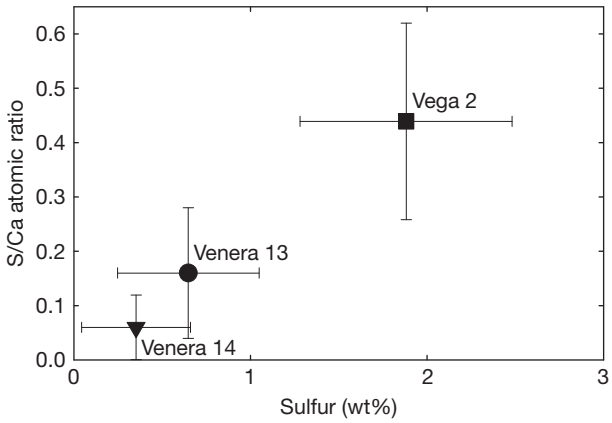


Figure 6 Sulfur content in Venus' rocks plotted against the S/Ca ratio. The S/Ca ratio in rocks is less than that in anhydrite (unity), indicating incomplete sulfurization of Ca-bearing minerals. Data are from Surkov et al. (1984, 1986).

On Venus, the instability of pyrrhotite and troilite with respect to oxidation (see **Section 10.12.3.1**) does not exclude the occurrence of pyrite, which can stably coexist with magnetite and hematite at the conditions of plains (**Figure 5**). The evaluations of mineral stability are limited by the uncertainties of thermodynamic data for gas and solid phases, impure solid phases, and uncertainties in the composition of the near-surface atmosphere. These factors led to controversial conclusions about pyrite stability on Venus (e.g., Barsukov et al., 1982; Fegley and Treiman, 1992; Fegley et al., 1997a; Klose et al., 1992; Wood, 1997).

Excess sulfur in surface materials on Venus and Mars likely has a volcanic source. On Mars, sulfur outgassing is consistent with the isotopic composition of S in Martian meteorites, which reveals photochemistry-driven mass-independent fractionation of sulfur in the atmosphere (Farquhar et al., 2000). On Venus, a volcanic origin of sulfur agrees with widespread volcanic features and elevated abundances of sulfur gases in the atmosphere. On both planets, volcanic and impact processes could have recycled sulfur through atmosphere and crustal reservoirs.

On Venus and Mars, the photochemical oxidation of sulfur gases produces sulfuric acid aerosols. In the sulfur-rich Venusian atmosphere, constantly produced aerosols form clouds at an altitude of 48–60 km that do not affect the surface (Esposito et al., 1997; Von Zahn et al., 1983). On Mars, sulfuric acid aerosols form after volcanic eruptions (Settle, 1979; Tian et al., 2010) and impact cratering (see **Section 10.12.3.8**). The negligible amounts of S-bearing gases in the present atmosphere (**Table 1**) suggest that they were consumed by weathering processes that followed the last volcanic or impact event. The deposition of sulfate aerosols leads to their neutralization through mineral alteration. As a result, sulfates form on the surfaces of mineral grains and rock fragments. In addition to sulfates, acid weathering by volcanic/impact aerosols on Mars would cause the deposition of silica and ferric oxides. Indeed, Fe^{3+}-rich coatings, likely nanophase oxides/oxyhydroxides, are ubiquitous on Martian surface materials (e.g., Goetz et al., 2005; Morris et al., 2006, 2008). Silica coatings/deposits have been suggested from orbital thermal infrared and near-infrared spectra (e.g., Michalski et al., 2005; Milliken et al., 2008; Roach et al., 2010) and detected by in situ investigations in Gusev crater (Ruff et al., 2011). The trapping of precipitated sulfuric acid aerosols in surface ices would have prevented the evaporation of aerosols and led to the acid-sulfate weathering of underlying rocks (Zolotov and Mironenko, 2007). The acid weathering of mineral dust in ices could have led to the formation of layered sulfate-rich rocks observed at the Meridiani Planum and other places (Niles and Michalski, 2009).

Although thermodynamics favor mineral weathering by sulfur gases under cold and dry Martian conditions (Gooding, 1978; Gooding et al., 1992), most proposed reactions remain poorly explored. Modeling experiments (Clark et al., 1979) indicate that conversions of carbonates and nitrates (if they are present) to corresponding sulfates are the

most likely pathways of interaction of gaseous sulfur oxides with surface minerals. However, low-temperature conversion of silicates to sulfates could be inhibited without aqueous phase.

10.12.3.3 Reactions with H_2O

Water vapor exists in the atmospheres of Mars and Venus (**Table 1**) and can interact with surface materials. On Mars, water condenses as ice, adsorbs on surface materials, and chemically bonds to minerals. On Venus, water vapor cannot condense anywhere near the surface and could be consumed via surface oxidation (e.g., reactions [3] and [8]) rather than through the formation of OH-bearing minerals. On both planets, early atmospheres could have contained a higher abundance of water (Donahue et al., 1997; Lammer et al., 2008; Owen, 1992), causing competitive hydration and oxidation of surface materials. Low-temperature conditions at elevation would have favored hydration and restricted dehydration. Many minerals may have lost water during subsequent evolution.

On Mars, the occurrence of water ice and hydrated minerals is controlled by their thermodynamic stability and kinetics of hydration and dehydration. Stabilities of ice and hydrated minerals on the surface vary with latitude and season. Ice and highly hydrated minerals (e.g., meridianite, $MgSO_4 \cdot 11H_2O$; antarcticite, $CaCl_2 \cdot 6H_2O$; and gypsum) are stable at high latitudes and during winter in midlatitudes. Elevated abundances of ground hydrogen at latitudes greater than ~40° inferred with neutron spectrometry from orbit (Boynton et al., 2002; Feldman et al., 2004; Mitrofanov et al., 2004) agree with thermodynamic models of ice stability (e.g., Mellon and Jakosky, 1995) and surface morphology features. Hydrated minerals may contribute to the high abundance of subsurface hydrogen (in ice + hydrates) at high latitudes. High H concentrations in some large equatorial regions (e.g., Arabia) are inconsistent with thermodynamic models of the latitudinal distribution of ice and hydrates. It is possible that highly hydrated minerals formed at low latitudes during periods of high obliquity (Kieffer and Zent, 1992) and have not dehydrated to minerals stable under current climatic conditions (Feldman et al., 2004). Slow dehydration of epsomite, $MgSO_4 \cdot 7H_2O$, and/or starkeyite, $MgSO_4 \cdot 4H_2O$ (e.g., Steiger et al., 2011; Wang et al., 2011), provides an explanation.

At Venus' surface, all hydrated minerals and the majority of OH-bearing minerals are unstable (Zolotov et al., 1997). Only some micas (e.g., phlogopite, $KMg_3Al_3Si_2O_{10}(OH,F)_2$) may be stable, especially at elevation. Partial substitution of the OH group by F favors phlogopite stability. Despite its stability, OH–F phlogopite formation requires transport of metals from different minerals, which is unlikely at present surface conditions. If micas and amphiboles were supplied from interior, they could survive at the current surface owing to slow dehydration (Johnson and Fegley, 2003, 2005). However, dehydration may have occurred at elevated atmospheric greenhouse temperatures caused by global volcanism 0.4 ± 0.1 Ga.

On both planets, the loss of water from the atmosphere in oxidation processes and H_2 escape could be compensated by internal processes. On Mars, polar caps, ground ice, and hydrated minerals are the major near-surface water reservoirs. Magmatic H_2O degassing could have played a role in earlier epochs. On Venus, atmospheric water could be resupplied via diffusion of H_2O and H_2 gases from nominally anhydrous minerals, which may account for the majority H_2O inventory (Zolotov et al., 1997). Volcanic degassing on Venus may not be an important source of atmospheric water because of suppressed explosive activity (Head and Wilson, 1986), a low abundance of H_2O in magmas, and its high solubility.

10.12.3.4 Reactions with CO_2

Carbon dioxide is the most abundant gas in the atmospheres of Mars and Venus that significantly affects mineral–gas interactions. On Venus, the massive CO_2 atmosphere accounts for the high surface pressure and temperature (see **Table 1**) and could be the major gas oxidizing surface materials (reactions [1] and [6]). On Mars, seasonal condensation and sublimation of CO_2 in polar regions are responsible for the annual cycle of atmospheric pressure (James et al., 1992). The presence of carbonates in Martian igneous rocks (meteorites), soils, and sedimentary deposits reveals the fixation of atmospheric CO_2 in the crust (e.g., Boynton et al., 2009; Ehlmann et al., 2008; Leshin et al., 2013; Morris et al., 2010).

At present conditions, a significant chemical fixation of CO_2 in carbonates is unlikely on both planets. Experimental data demonstrate that carbonatization of silicates is favored under aqueous conditions (Tanner et al., 1985), which are absent at current surfaces. Although the formation of Martian carbonates on irradiated surfaces has been predicted in theoretical models (Huguenin, 1974, 1976), an experimental work (Mukhin et al., 1996) reveals the possibility of photostimulated decomposition of carbonates. Booth and Kieffer (1978) experimentally demonstrated the formation of only traces of carbonates during the interaction of Ca silicates with CO_2 under Martian temperatures with or without mild irradiation. It follows that carbonates observed in Martian materials could have formed in aqueous environments.

Although low totals in Venus' surface analysis (**Table 2**) may reveal the presence of carbonates (Fegley et al., 1997a), the formation of calcite ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$) is forbidden because of their thermodynamic instability with respect to high fugacity of atmospheric SO_2 (e.g., Fegley and Treiman, 1992; Volkov et al., 1986), which would destroy them rapidly (Fegley and Prinn, 1989). If indigenously formed calcite is present on Venus' surface, the conversion of calcite to anhydrite may occur via net reaction

$$CaCO_3 + 1.5SO_2 \rightarrow CaSO_4 + CO_2 + 0.25S_2 \qquad [13]$$

which was studied by Terradellas and Bonnetain (1973), and would result in rapid destruction of calcite at Venus' surface temperature. Siderite ($FeCO_3$) is unstable with respect to oxidation to ferric oxides. On the Martian surface, Ca carbonates could be attacked by SO_2, SO_3 (Clark et al., 1979; Gooding et al., 1992), acidic aerosols and rains of volcanic/impact origin, and other low-pH fluids, as illustrated by the reactions

$$CaCO_3 + SO_2 + 0.5O_2 \rightarrow CaSO_4 + CO_2 \qquad [14]$$

$$CaCO_3 + 2H^+ + SO_4^{2-} + H_2O \rightarrow CaSO_4 \cdot 2H_2O + CO_2 \quad [15]$$

$$CaCO_3 + 2H^+ + 2Cl^- + 5H_2O \rightarrow CaCl_2 \cdot 6H_2O + CO_2 \quad [16]$$

Magnesite ($MgCO_3$) is the only rock-forming carbonate stable with respect to SO_2 on the surfaces of Venus and Mars. Its occurrence at Venus' surface is unlikely because of thermodynamic and kinetic problems for its formation at present conditions. On Mars, a detection of Mg-rich carbonates in some surface materials (Ehlmann et al., 2008; Morris et al., 2010) and possibly in atmospheric dust (Bandfield et al., 2003) and soils (Leshin et al., 2013) may in part reflect their relative stability with respect to S oxides and sulfuric acid solutions.

The rapid destruction of calcite and dolomite compared to slow, if any, carbonatization of surface materials implies a likely absence of carbonates at the surface of Venus. An exposure of ancient sedimentary carbonates is not consistent with basaltic lava morphology of Venus' plains, and long sinuous channels (Baker et al., 1997) could have been formed by molten chlorides (see **Section 10.12.3.5**) rather than low-viscosity carbonate magmas. The paucity of carbonates on Venus agrees with the fact that there are similar masses of CO_2 in Venus' atmosphere and Earth's carbonates (**Table 4**). It follows that a popular idea of buffering atmospheric CO_2 by the Urey (1952) carbonate–silicate equilibrium,

$$CaCO_3 + SiO_2 \rightarrow CaSiO_3 + CO_2 \quad [17]$$

(e.g., Adamchik and Draper, 1963; Fegley and Treiman, 1992; Lewis, 1970; Mueller, 1964), may be invalid despite the fact that the equilibrium PCO_2 in reaction [17] at 740 K is similar to the PCO_2 at Venus' surface. In addition, equilibrium [17] could be unstable with respect to variations in atmospheric temperature and/or pressure [Hashimoto et al., 1997; Treiman and Bullock, 2011]. However, a possibility for equilibrium [17] may exist below the surface (see **Section 10.12.5**).

On Mars, secondary carbonates in crustal igneous rocks and sedimentary deposits may account for the major reservoir of CO_2 formed in aqueous conditions. It is possible that carbonates were deposited in a primordial ocean on Venus, if it ever existed. Subsequent volcanic and impact degassing of chemically active gases would have led to chemical weathering of surface carbonates on both planets. Sulfates and chlorides are the likely minerals formed after carbonates (reactions [13]–[16]). On Mars, subsurface carbonates may not be much affected, except for local dissolution by acidic fluids of near-surface origin. On Venus, early formed carbonates (if they formed) could be partially/completely decomposed, contributing to greenhouse heating from a thickening atmosphere (see **Section 10.12.5**).

10.12.3.5 Reactions with Halogen-Bearing Species

Halogens are delivered into planetary atmospheres through magmatic and impact degassing. On Earth (Oppenheimer, 2003; Symonds et al., 1994) and Mars, they degas as hydrogen halides (HCl and HF) and SiF_4. On Venus, hydrogen-depleted magmas can emit diatomic halides, S- and S–O-halides, and SiF_4. The low abundances of Cl- and F-bearing gases in the atmospheres of Mars and Venus (**Table 1**) imply that degassed halogens have been consumed by surface materials. It is likely that planetary crusts are enriched in previously degassed halogens that are mainly stored in metal halides (e.g., NaCl, KCl, $CaCl_2$, and $MgCl_2$) and halogen-bearing salts (e.g., apatite). Chlorides could be the most abundant halogen compounds owing to the higher solar abundance of Cl compared with other halogens. In addition, significant quantities of fluorine could be preserved in igneous rocks, especially those with elevated alkali content (Venera 8 and Venera 13 landing sites and Columbia Hills basalts). On both planets, heating of surfaces and crusts could have caused the mobilization, redistribution, and recycling of halogens. These processes may include magmatic/impact degassing, assimilation in magma chambers, melting and release of halogen-bearing salts as well as aqueous activity, and brine formation and evaporation.

Martian surface materials are characterized by elevated and variable Cl and Br contents compared to basaltic rocks. The incorporation of degassed halogens in the soil could have occurred through precipitation of volcanic/impact aerosols, dissolution in aqueous solutions, and gas–mineral-type reactions. The hydrolysis of hydrogen halides produces acidic solutions that dissolve surface minerals followed by precipitation of halides of Ca, Mg, Na, K, and Fe upon freezing or evaporation. Weathering of rocks by acid gases, aerosols, acid rains, and surface solutions would convert primary and secondary minerals to corresponding halides. These transformations can be illustrated by reactions [16] and [18]:

$$Ca_2Si_2O_6(\text{in pyroxene}) + 4HCl(\text{gas, aerosol, solute}) \rightarrow 2CaCl_2 \cdot 2H_2O + 2SiO_2 \quad [18]$$

Throughout Martian history, these highly water-soluble salts were probably mobilized and redeposited by sporadic aqueous solutions, consistent with the occurrence of chloride-rich deposits in Martian highlands (Osterloo et al., 2010). At present, deep crustal brines are most likely enriched in Cl and Br, which lower the freezing temperature (Burt and Knauth, 2002). Variable S/Cl/Br ratios in layered deposits at Meridiani Planum are consistent with element separation in aqueous processes (Clark et al., 2005; Marion et al., 2009). Surface materials also contain perchlorates likely formed through photochemical oxidation of chlorides in the atmosphere and/or at the surface (Catling et al., 2010; Leshin et al., 2013).

Hydrogen halides (HCl and HF) are present in trace amounts in Venus' atmosphere (**Table 1**). These gases could have originated from hydrolysis of H-free volcanic gases (e.g., Cl_2, SCl_2, and SiF_4) and halogen-bearing surface materials (apatite and glasses). In turn, their interactions with primary and secondary minerals would lead to the formation of halogen salts. Although halogens have not been analyzed in Venus' rocks, the high upper limits of Cl (**Table 2**) do not exclude elevated abundances of halogens.

It is possible that atmospheric abundances of HCl and HF reflect chemical equilibria among these gases and secondary minerals at the surface (Fegley and Treiman, 1992; Lewis, 1970; Mueller, 1968). The low masses of atmospheric halogens (**Table 4**) compared with possible masses of halogens in a permeable surface layer imply that the atmospheric abundances of halogens are controlled by surface mineralogy.

Atmospheric masses of Cl and F can be consumed in global layers 22 and 0.15 cm thick, respectively (at 0.1 wt% of Cl and F). A comparison of the mass of Cl in Venus' atmosphere with that in terrestrial oceans and sediments (**Table 4**) suggests a comparable reservoir of halogens in Venus' crust (Fegley and Treiman, 1992; Zolotov and Volkov, 1992).

Venus' halides could have been mobilized in the crust because of their low melting points compared to silicates. Channels (Baker et al., 1997) and crater outflows observed in Magellan radar images may have been formed by molten halides. Some halides could have accumulated in the crust through evaporation of putative early ocean. It is possible that eutectic melts of halogen salts currently exist in Venus' crust. Likewise, Cl–Br–Ca–Mg–Na–(SO_4) eutectic aqueous brines may be present deep below the Martian surface.

10.12.3.6 Condensation of Metal and Nonmetal Species: The Role of Altitude

Volcanism, impacts, gas–solid-type weathering, and thermal rock outgassing delivered many elements into planetary atmospheres before they condensed. Terrestrial silicate volcanism degasses many metals (e.g., Na, K, Bi, Hg, Pb, Cu, Zn, Cd, and Sb) and nonmetals (B, Se, and As) (Oppenheimer, 2003). On Venus and Mars, volcanic sources of S in surface materials may imply elevated abundances of Se, Te, As, and volatile metals.

A sizeable temperature gradient between Venus' plains (730–750 K) and mountaintops (650–700 K) should cause preferential condensation of heavy elements in elevated regions. Thermodynamic calculations demonstrate that sulfides, sulfosalts, halides, and chalcogenides of Pb, Bi, As, Cu, and Sb may condense on mountaintops and account for the high dielectric constant (Brackett et al., 1995; Schaefer and Fegley, 2004). Lower temperatures at elevation lead to lower rates of chemical weathering and would favor longer preservation of locally unstable phases. It follows that some metal/sulfide condensates may be unstable with respect to gas–solid weathering reactions.

10.12.3.7 Kinetic Aspects of Surface–Gas-Type Reactions

Gas–solid-type reactions affect individual mineral grains that are subjected to isochemical weathering. As a result, alteration products should consist of mechanical mixtures of minerals formed through weathering of primary phases. Because weathering products form at the surface of mineral grains, smaller particles are subjected to more alteration and secondary particles are smaller than initial grains. In many cases, weathering is controlled by diffusion of reacting gases through a layer of alteration products. Physical removal of alteration products from mineral surfaces would increase the rate of chemical weathering. It follows that small particles could be most chemically altered, consistent with observations on Mars (e.g., Yen et al., 2005) and Venus (Shkuratov et al., 1987). Despite wind abrasion and thermodynamic potential for weathering reactions at planetary surfaces, many gas–mineral-type interactions are inefficient because of slow reaction rates, especially at the diffusion-controlled stage. As a result, significant portions of surface rocks and fine-grained materials on Mars and Venus remain chemically unaltered.

On Venus, kinetic parameters of key oxidation reactions [1]–[3], [6]–[8] and sulfurization reactions [11] and [12] remain unknown. Although rapid formation of anhydrite through SO_2–calcite interaction is evident from many experiments (e.g., Fegley and Prinn, 1989; Terradellas and Bonnetain, 1973), this reaction may not be relevant to the present-day Venus' surface (see **Section 10.12.3.4**).

On Mars, rates and mechanisms of gas–solid weathering processes remain poorly explored, partially because of low reaction rates at subzero temperatures and low densities of atmospheric gases. The formation and reactivity of carbonates through gas–solid reactions (Booth and Kieffer, 1978; Mukhin et al., 1996) remain to be confirmed, including the effects of irradiation and gas diffusion through reaction products. The study of photostimulated oxidation, hydration (e.g., Huguenin, 1973a,b, 1974, 1976, 1982; Morris and Lauer, 1980), and dehydration (Morris and Lauer, 1981; Yen et al., 1999) could be revisited. Despite modeling experiments by Clark et al. (1979), kinetic parameters of low-temperature weathering by sulfur and hydrogen halide gases remain unknown. Recent experimental investigations of hydration and dehydration of sulfates could be used to evaluate timing of H_2O exchange between the atmosphere and hydrated salts (e.g., Steiger et al., 2011; Wang et al., 2011).

10.12.3.8 Local Reactions Caused by Impacts and Volcanic Eruptions

Impacts and volcanic eruptions create highly variable local environments in which chemical transformations are governed by mass transfer and reaction rates. In impact events, thermal dissociation of species leads to chemically contrasting gas mixtures (e.g., H_2, O_2, CO, CO_2, SO_3, and nitrogen oxides) that affect local and global atmospheric environments (e.g., Kawaragi et al., 2009; Pope et al., 1997; Prinn and Fegley, 1987; Toon et al., 1997). Reactions in volcanic/impact sites can include the oxidation of ferrous iron and sulfides, reduction of iron to Fe^0 metal at $T > 1500$ K, dehydration and hydration of minerals, thermal decomposition of carbonates, and conversion of carbonates to sulfates. On Earth, oxidation by volcanic gases mixed with air often leads to hematite-bearing pumice with a reddish color. On Mars and Venus, high-temperature oxidation by H_2O and/or CO_2 in impact/volcanic gases could have contributed to the formation of sulfates and ferric oxides in soil and dust. Although hot CO_2 could be the major oxidizing agent on Venus (Fegley et al., 1995), impacts of comets and carbonaceous chondrites may cause some oxidation by H_2O. On both planets, hematite could have formed via oxidation in impact/volcanic events.

On Mars, slow cooling of lavas and impact melts from inefficient convection and low atmospheric density allowed a great degree of chemical alteration of melts and hot rocks. However, the low PO_2 in the atmosphere, especially during periods of volcanic activity, would not have favored rapid oxidation of hot surface materials. On Venus, rapid cooling of silicate melts at the contact with the dense atmosphere (Frenkel and Zabalueva, 1983) could have been compensated by the high densities of chemically active gases (CO_2 and SO_2).

On Mars, volcanic/impact melting of ice-rich deposits could have led to local hydration, especially in fluidized

outflow deposits. Materials in Martian surge deposits could have been modified (oxidized, hydrated, and weathered) shortly after deposition (Knauth et al., 2005). On Venus, an impact-generated layered material suggested at Venera landing sites (e.g., Basilevsky et al., 2004) may be composed of materials that were altered by impact events. The presence of chemically distinct fine-grained material in parabolic crater ejecta deposits is consistent with their elevated dielectric constant (up to eight) (Campbell et al., 1992; Pettengill et al., 1997).

On Earth, Mars, and Venus, impact/volcanic degassing of sulfur-rich rocks would lead to an array of sulfur species and could affect global atmospheric composition. The impact decomposition of crustal sulfates during the Chicxulub event at the Cretaceous–Tertiary boundary released sulfur oxides into the atmosphere followed by downpours of acid rain (Gupta et al., 2001; Pope et al., 1997). On sulfate-rich Mars, raining from impact-generated clouds (Segura et al., 2002, 2012) could be responsible for observed signs of acid-sulfate weathering (Zolotov and Mironenko, 2007). The thermal decomposition of carbonates during impacts and/or their interaction with released sulfur oxides and acid aerosols/rainfalls would have altered and/or obliterated surface carbonates on Mars. On Venus, impact/volcanic release of sulfur gases would temporarily affect the mass of sulfuric acid clouds.

10.12.4 Wind-Related Processes

Eolian processes provide the major mechanism for mass transfer on contemporary surfaces of Mars and Venus (Greeley and Iversen, 1985). Both erosional (yardangs) and depositional (dunes) features are observed (Greeley et al., 1992, 1997). On Mars, dust devils and global dust storms are widely observed from orbit and the surface (Balme and Greeley, 2006). On both planets, the orientation of eolian features is consistent with general circulation models.

The effectiveness of eolian transport depends on air density, temperature, and gravity, which are significantly different on Mars and Venus (Greeley et al., 1992, 1997). Nevertheless, models and experiments demonstrate that particles that are a few hundred microns in diameter can be transported on both planets. Particles with a diameter of 100 μm represent the most mobile fractions (Greeley et al., 1980, 1992, 1997). Larger and smaller particles require stronger winds to be transported.

On Mars, low gravity and high wind velocity overcome the effects of low air density and cause significant eolian transport. Dust with grain sizes of a few microns is homogenized in the atmosphere on a planetary scale. Globally, deflation characterizes some low-albedo regions and accumulation is typical in high-albedo regions. Larger particles (up to perhaps a centimeter in diameter) are transported by saltation and creep, forming areas dominated by deflation and accumulation. Martian polar deposits (Thomas et al., 1992) and layered rocks with cross-beddings investigated by MER (e.g., Grotzinger et al., 2005) indicate a substantial eolian transport and accumulation throughout history.

On Venus, measurements of wind speed near the surface (<2 m s^{-1}, Ksanfomality et al., 1993), extrapolation of wind velocities to the surface (0.5–1 m s^{-1}, Counselman et al., 1979), and wind tunnel experiments (Greeley et al., 1987) suggest the possibility of present-day eolian activity of Venus. This is consistent with the observation of wind-related features (mostly wind streaks) in Magellan radar images (Greeley et al., 1997). Fine-grained material is mainly supplied from impact cratering and chemical weathering. The apparent paucity of dunes on Venus could be related to a lack of sand-size particles.

The orientation of Venusian wind features indicates the upper altitude superrotation of the atmosphere and the extension of Hadley circulation to the surface (Greeley et al., 1997). Wind transport is consistent with the presence of parabola-shaped zones of low radar backscatter around the relatively fresh impact craters. Models show that wind activity would disperse the inferred fine-grained materials in the parabolas over a few tens of millions years.

Without aqueous solutions, physical weathering of surface rocks is mostly restricted to wind abrasion and rock disintegration caused by chemical weathering of individual grains. Despite vivid signs of eolian activity, the estimated rates of eolian erosion are small on both planets. On Mars, an erosion rate of ~0.03 nm year^{-1} indicates slight surface changes since the Hesperian epoch (~3 Ga) (Golombek et al., 2006). Modeling experiments demonstrate the possibility of rock abrasion even for moderate wind velocities (0.5–0.7 m s^{-1}) at the surface of Venus (Greeley et al., 1987). Despite these results, volcanic features remain well preserved (Crumpler et al., 1997). In particular, during the Magellan mission, no surface changes were observed that could be attributed to eolian processes.

10.12.5 Atmosphere–Surface Interactions Throughout History

Although gas–solid-type interactions prevailed over the majority of planetary histories, the effectiveness of these interactions varied over time along with changes in surface temperatures and atmospheric compositions. On Mars and Venus, significant atmosphere–surface interaction could have occurred earlier in history under wetter conditions and different temperatures, as well as through multiple episodes of volcanic eruptions and impacts.

Elevated D/H ratios in the current atmospheres of Mars ($7.8 \pm 0.3 \times 10^{-4}$) and Venus ($2.4 \pm 0.5 \times 10^{-2}$) compared to the Earth's oceanic value (1.56×10^{-4}) indicate more humid conditions in their earlier atmospheres, which deteriorated by the escape of H (Donahue et al., 1997; Krasnopolsky and Feldman, 2001; Owen, 1992). On Mars, a low $^{14}N/^{15}N$ ratio in the atmosphere (170×15; Owen, 1992) compared to the Earth's value (272) implies gas escape from a denser early atmosphere. Despite a low solar luminosity, the greenhouse effect in denser and/or more humid atmospheres could have produced elevated surface temperatures on Mars and Venus. However, that cloud condensation should have affected gas abundances (e.g., CO_2 condensation on Mars, Kasting (1991)) and surface temperatures through albedo changes (Hashimoto and Abe, 2001).

The oxidation of ferrous minerals and sulfides by water (reactions [3] and [8]) and H_2 production is a likely process on early planets (Lewis and Prinn, 1984; Wänke and Dreibus, 1987). H_2 escape contributes to net oxidation of surfaces and

atmospheres. The high oxidation state of most Martian meteorites is consistent with early oxidation of the lithosphere by H_2O.

Hydration of surface material is another likely process on early planets. On Mars, somewhat warmer and wetter conditions during the Noachian epoch are supported by the presence of phyllosilicates (Carter et al., 2013; Ehlmann et al., 2011; Murchie et al., 2009; Poulet et al., 2005), valley networks (Carr, 1996), and heavily eroded terrains (Craddock and Howard, 2002). On Venus, the evolution of the greenhouse effect is poorly constrained and early hydration remains conjectured. The majority of hydrated phases would have been dehydrated during subsequent evolution, especially during increases in surface temperatures.

Humid conditions in the early atmospheres would have favored the fixation of atmospheric CO_2 in carbonates. On Mars, the formation of carbonates in aqueous environments provides a major pathway to diminish an early CO_2 atmosphere. On Venus, carbonates could have been deposited in a primordial ocean, if one ever existed. Subsequent metamorphic reactions in the lithosphere could have led to the establishment of carbonate–silicate equilibria below the surface. Such a mechanism might be consistent with a buffering atmospheric CO_2 by equilibrium [17]. Calcium-bearing minerals below the surface should be protected from atmospheric SO_2, HCl, and HF. These gases could be consumed through weathering in upper subsurface layers and may not reach putative carbonate-bearing rocks at depth.

Throughout the geologic histories of Mars and Venus, volcanic activity and impact events would have affected atmospheric composition and surface temperature (Fanale et al., 1992). In particular, injection of water and CO_2 into the atmosphere would have increased the surface temperature through greenhouse heating. On Mars, elevated temperatures would have led to further supplies of H_2O and CO_2 from polar caps and crustal reservoirs, which may have stabilized liquid water on the surface.

On Venus, intensive volcanism 0.4 ± 0.1 Ga ago (Basilevsky et al., 1997) would have been accompanied by IR-absorbing volcanic gases and a corresponding increase in surface temperature. Even a small fraction of water vapor in the products of volcanic degassing would have led to significant rising of atmospheric and lithospheric temperature. The temperature rise would have facilitated partial melting of igneous rocks and softening of the lithosphere and triggered further volcanism (Solomon et al., 1999). High-temperature reactions of atmospheric gases with hot surface materials could have led to significant alteration of surface rocks. In particular, degassed SO_2 and halogens could have been partially consumed by surface materials. Volcanically released H_2O would have been consumed by high-temperature oxidation of rocks (reactions [3] and [8]). It is possible that most chemical alterations currently observed on the surface occurred during the short but intense period of volcanic activity. Since the resurfacing event, abundances of minor atmospheric gases and the redox state may have been controlled by gas–mineral reactions. Changes in surface temperature altered the partial pressures of gases in gas–mineral-type reactions. For example, the oxidation state of the atmosphere and abundances of S-bearing gases could have been controlled by surface reactions that involve magnetite, hematite, and pyrite.

10.12.6 Summary and Unsolved Questions

Despite significant differences in environmental conditions on Mars and Venus, atmosphere–surface interactions reveal several common features. On both planets, atmospheric gases have oxidized surface materials. The dissociation of H_2O and hydrogen escape from early atmospheres could have caused partial oxidation of crustal materials. Although early atmospheric conditions could have favored the formation of carbonates and hydrated minerals, current environments may not be feasible for carbonate formation. Throughout history, volcanic/impact degassing was followed by consumption of sulfur and halogens into surface materials through weathering reactions. On both planets, volcanic/impact degassing, surface temperature, and gas–mineral reactions are interrelated through the greenhouse heating and cloud feedbacks.

Specifics of atmosphere–surface interaction on Mars include (1) seasonal cycles of condensation and sublimation of H_2O and CO_2, (2) possible seasonal hydration and dehydration of some salts, (3) atmosphere–surface redistribution of H_2O and CO_2 and hydration and dehydration of minerals during obliquity cycles, (4) low-temperature inhibition of a majority of thermochemical chemical reactions at the present surface, (5) suppressed crystallization of amorphous materials, (6) a possible contribution of photostimulated reactions involving minerals and organic compounds, (7) a possible liquid water–rock reactions below the surface, and (8) potential formation of the majority of secondary minerals by participation of aqueous solutions earlier in its history.

Venus' specifics include (1) oxidation and sulfurization of surface rocks through gas–solid-type reactions, (2) isochemical weathering of individual solid phases with respect to nonvolatile elements at Venus' surface temperature (e.g., Al, Si, Mg, Fe, Ca, and Na), (3) unlikely current hydration and a possibility of dehydration of early formed phases, (4) the possibility of gas–gas and gas–mineral equilibria at the atmosphere–surface interface that determine the oxidation state and sulfur and halogen contents in the atmosphere, (5) a strong altitudinal effect for chemistry and physics of gas–surface interactions, (6) chemical weathering of minerals by unequilibrated atmospheric gases at elevations, (7) possible condensation of heavy metal compounds on mountaintops, and (8) the possibility for efficient gas–mineral-type reactions during intensive global volcanism.

A few of unsolved questions are the following: Under what environmental conditions were planetary surface materials oxidized and enriched in volatiles? What is the contribution of gas–mineral-type reactions to the composition of planetary surfaces? How have gas–mineral interactions changed over planetary history? What is the effect of irradiation on mineralogical transformations on the Martian surface? What are the abundances of C, halogens, and H in surface rocks on Venus? Do Venus' atmospheric gases equilibrate with each other and surface minerals? Which atmospheric gases are controlled by atmosphere–surface reactions on Venus? What are the alteration pathways and secondary mineralogies on Venus' lowlands and highlands?

Further progress could be made by integration of space missions, telescopic observations, experimental work, and theoretical models. A mission to Venus should aim to determine the composition of the near-surface atmosphere and the mineralogy of the surface. Experimental efforts focused on the kinetics of gas–mineral reactions on Venus and Mars should be complimented with coupled kinetic–thermodynamic models.

References

Adamchik JA and Draper AL (1963) The temperature dependence of the Urey equilibrium and the problem of CO_2 content of the atmosphere of Venus. *Planetary and Space Science* 11: 1303–1307.

Atreya SK, Wong AS, Renno NO, et al. (2006) Oxidant enhancement in Martian dust devils and storms: Implication for life and habitability. Space physics, Mars and life. *Astrobiology* 6(3): 439–450.

Avduevsky VS, Godnev AG, Zakharov YV, et al. (1983) An estimate of the physical and mechanical characteristics of the soil of Venus from measurements of the impact overload during the landing of the Venera 13 and Venera 14 automatic interplanetary stations. *Cosmic Research* 21: 260–268.

Baker VR, Komatsu G, Gulick VC, and Parker TJ (1997) Channels and valleys. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 757–796. Tucson, AZ: University of Arizona Press.

Balme M and Greeley R (2006) Dust devils on Earth and Mars. *Reviews of Geophysics* 44, RG3003.

Bandfield JL, Glotch TD, and Christensen PR (2003) Spectroscopic identification of carbonate minerals in the Martian dust. *Science* 301: 1084–1087.

Banin A, Clark BC, and Wänke H (1992) Surface chemistry and mineralogy. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 594–625. Tucson, AZ: University of Arizona Press.

Barmin IV and Shevchenko AA (1983) Soil-scooping mechanism for the Venera 13 and Venera 14 unmanned interplanetary spacecraft. *Cosmic Research* 21: 118–121.

Barsukov VL, Bazilevskii AT, Burba GA, Bobina NN, and Kriuchkov VP (1986) The geology and geomorphology of the Venus surface as revealed by the radar images obtained by Veneras 15 and 16. *Journal of Geophysical Research: Proceedings of the 16th LPSC* 91: D378–D398.

Barsukov VL, Volkov VP, and Khodakovsky IL (1982) The crust of Venus: Theoretical models of chemical and mineral composition. *Journal of Geophysical Research: Proceedings of the 13th LPSC* 87: A3–A9.

Basilevsky AT, Head JW, and Abdrakhimov AM (2004) Impact crater air fall deposits on the surface of Venus: Areal distribution, estimated thickness, recognition in surface panoramas, and implications for provenance of sampled surface materials. *Journal of Geophysical Research* 109: E12003.

Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1047–1086. Tucson, AZ: University of Arizona Press.

Basilevsky AT, Kuzmin RO, Nikolaeva OV, et al. (1985) The surface of Venus as revealed by the Venera landings: Part II. *Geological Society of America Bulletin* 96: 137–144.

Bell JF, McSween HY, Crisp JA, et al. (2000) Mineralogic and compositional properties of Martian soil and dust: Results from Mars pathfinder. *Journal of Geophysical Research* 105: 1721–1755.

Blake DF, Morris RV, Kocurek G, et al. (2013) Curiosity at Gale crater, Mars: Characterization and analysis of the Rocknest sand shadow. *Science* 341: 1239505.

Booth MC and Kieffer HH (1978) Carbonate formation in Marslike environments. *Journal of Geophysical Research* 83: 1809–1815.

Boynton WV, Feldman WC, Squyres SW, et al. (2002) Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* 297: 81–85.

Boynton WV, Ming DW, Kounaves SP, Young SM, Arvidson AE, Hecht MH, et al. (2009) Evidence for calcium carbonate at the Mars Phoenix landing site. *Science* 325: 61–64.

Brackett RA, Fegley B, and Arvidson RE (1995) Volatile transport on Venus and implications for surface geochemistry and geology. *Journal of Geophysical Research* 100: 1553–1563.

Bridges JC, Catling DC, Saxton JM, Swindle TD, Lyon IC, and Grady MM (2001) Alteration assemblages in Martian meteorites: Implications for near-surface processes. *Space Science Reviews* 96(1/4): 365–392.

Burt DR and Knauth LP (2002) Electrically conducting, Ca-rich brines, rather than water, expected in the Martian subsurface. *Journal of Geophysical Research* 108: 8026. http://dx.doi.org/10.1029/2002JE001862.

Campbell DB, Stacy NJS, Newman WI, et al. (1992) Magellan observations of extended impact-crater related features on the surface of Venus. *Journal of Geophysical Research* 97: 16249–16278.

Carr MH (1996) *Water on Mars*. New York: Oxford University Press, 229 pp.

Carter J, Poulet F, Bibring J-P, Mangold N, and Murchie SL (2013) Hydrous minerals on Mars as seen by the CRISM and OMEGA imaging spectrometers: Updated global view. *Journal of Geophysical Research: Planets* 118: 831–858. http://dx.doi.org/10.1029/2012JE004145.

Catling DC, Claire MW, Zahnle KJ, et al. (2010) Atmospheric origins of perchlorate on Mars and in the Atacama. *Journal of Geophysical Research* 115: E00E11.

Christensen PR and Moore HJ (1992) The Martian surface layer. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 686–729. Tucson, AZ: University of Arizona Press.

Christensen PR, Morris RV, Lane MD, Bandfield JL, and Malin MC (2001) Global mapping of Martian hematite deposits: Remnants of water-driven processes on early Mars. *Journal of Geophysical Research* 106: 23873–23886.

Clark BC, Baird AK, Weldon RJ, Tsusaki DM, Schabel L, and Candelaria MP (1982) Chemical composition of Martian fines. *Journal of Geophysical Research* 87: 10059–10067.

Clark BC, Kenley SL, O'Brien DL, Huss GR, Mack R, and Baird AK (1979) Heterogeneous phase reactions of Martian volatiles with putative regolith materials. *Journal of Molecular Evolution* 14: 91–102.

Clark BC, Morris RV, McLennan SM, et al. (2005) Chemistry and mineralogy of outcrops at Meridiani Planum. *Earth and Planetary Science Letters* 240: 73–94.

Counselman CC, Gourevitch SA, King RW, Loriot GB, and Prinn RG (1979) Venus winds are zonal and retrograde below the clouds. *Science* 205: 85–87.

Craddock RA and Howard AD (2002) The case for rainfall on a warm, wet early Mars. *Journal of Geophysical Research* 107(E11): 5111.

Crumpler LS, Aubele JC, Senske DA, Keddie ST, Magee KP, and Head JW (1997) Volcanoes and centers of volcanism on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 697–756. Tucson, AZ: University of Arizona Press.

Delory GT, Farrell WM, Atreya SK, et al. (2006) Oxidant enhancement in dust devils and storms: Storm electric fields and electron dissociative attachment. *Astrobiology* 6: 451–462.

Donahue TM, Grinspoon DH, Hartle RE, and Hodges RR Jr. (1997) Ion/neutral escape of hydrogen and deuterium: Evolution of water. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 385–414. Tucson, AZ: University of Arizona Press.

Ehlmann BL, Mustard JF, Murchie SL, et al. (2008) Orbital identification of carbonate-bearing rocks on Mars. *Science* 322: 1828–1832.

Ehlmann BL, Mustard JF, Murchie SL, et al. (2011) Subsurface water and clay mineral formation during the early history of Mars. *Nature* 479: 53–60.

Ekonomov AP, Golovin YuM, and Moshkin BE (1980) Visible radiation observed near the surface of Venus – Results and their interpretation. *Icarus* 41: 65–75.

Encrenaz T, Greathouse TK, Lefévre F, and Atreya SK (2012) Hydrogen peroxide on Mars: Observations, interpretation and future plans. *Planetary and Space Science* 68: 3–17.

Esposito LW, Bertaux J-L, Krasnopolsky V, Moroz VI, and Zasova LV (1997) Chemistry of lower atmosphere and clouds. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 415–458. Tucson, AZ: University of Arizona Press.

Fanale FP, Postawko SE, Polack JB, Carr MH, and Pepin RO (1992) Mars: Epochal climate change and volatile history. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1135–1179. Tucson, AZ: University of Arizona Press.

Farquhar J, Savarino J, Jackson TL, and Thiemens MH (2000) Evidence of atmospheric sulphur in the Martian regolith from sulphur isotopes in meteorites. *Nature* 404: 50–52.

Fegley B Jr. (2014) Venus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 1, chapter 2.7. Oxford: Elsevier, 2nd edn.

Fegley B Jr., Klingelhöefer G, Brackett RA, Izenberg N, Kremser DT, and Lodders K (1995) Basalt oxidation and the formation of hematite on the surface of Venus. *Icarus* 118: 373–383.

Fegley B Jr., Klingelhöefer G, Lodders K, and Widemann T (1997a) Geochemistry of surface–atmosphere interactions on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 591–636. Tucson, AZ: University of Arizona Press.

Fegley B Jr. and Prinn RG (1989) Estimation of the rate of volcanism on Venus from reaction rate measurements. *Nature* 337: 55–58.

Fegley B and Treiman AH (1992) Chemistry of atmosphere–surface interactions on Venus and Mars. In: Luhmann JG, Tatrallyay M, and Pepin RO (eds.) *Venus and Mars: Atmospheres, Ionospheres, and Solar Wind Interactions*. *Geophysical Monograph*, vol. 66, pp. 7–71. Washington, DC: American Geophysical Union.

Fegley B Jr., Zolotov MYu, and Lodders K (1997b) The oxidation state of the lower atmosphere and surface of Venus. *Icarus* 125: 416–439.

Feldman WC, Prettyman TH, Maurice S, et al. (2004) Global distribution of near-surface hydrogen on Mars. *Journal of Geophysical Research* 109: E09006.
Fergason RL, Christensen PR, Bell JF III, et al. (2006) Physical properties of the Mars exploration rover landing sites as inferred from Mini-TES-derived thermal inertia. *Journal of Geophysical Research* 111: E02S21.
Florensky CP, Bazilevakii AT, Kriuchkov VP, et al. (1983a) Venera 13 and Venera 14 – Sedimentary rocks on Venus? *Science* 221: 57–59.
Florensky CP, Nikolaeva OV, Volkov VP, et al. (1983b) Redox indicator 'CONTRAST' on the surface of Venus. *Lunar and Planetary Science* XIV: 203–204 (abstract, The Lunar and Planetary Institute, Houston).
Florensky CP, Ronca LB, Basilevsky AT, et al. (1977) The surface of Venus as revealed by Soviet Venera 9 and 10. *Geological Society of America Bulletin* 88: 1537–1545.
Ford PG and Pettengill GH (1983) Venus: Global surface radio emissivity. *Science* 220: 1379–1381.
Frenkel MYa and Zabalueva EV (1983) The solidification of effusive magmas on Venus and Earth. *Geochemistry International* 20(5): 43–47.
Garvin JB, Head JW, Zuber MT, and Helfenstein P (1984) Venus: The nature of the surface from Venus panoramas. *Journal of Geophysical Research* 89: 3381–3399.
Gellert R, Rieder R, Brückner J, et al. (2006) Alpha particle X-ray spectrometer (APXS): Results from Gusev crater and calibration report. *Journal of Geophysical Research* 111: E02S05.
Gendrin A, Mangold N, Bibring J-P, et al. (2005) Sulfates in Martian layered terrains: The OMEGA/Mars express view. *Science* 307: 1587–1591.
Glotch TD, Bandfield JL, Christensen PR, et al. (2006) Mineralogy of the light-toned outcrop at Meridiani Planum as seen by the miniature thermal emission spectrometer and implications for its formation. *Journal of Geophysical Research* 111: E12S03.
Goetz W, Bertelsen P, Binau CS, et al. (2005) Indication of drier periods on Mars from the chemistry and mineralogy of atmospheric dust. *Nature* 436: 62–65.
Golombek MP, Grant JA, Crumpler LS, et al. (2006) Erosion rates at the Mars Exploration Rover landing sites and long-term climate change on Mars. *Journal of Geophysical Research* 111: E12S10.
Gooding JL (1978) Chemical weathering on Mars: Thermodynamic stability of primary minerals (and their alteration products) from mafic igneous rock. *Icarus* 33: 483–513.
Gooding JL (1992) Soil mineralogy and chemistry on Mars: Possible clues from salts and clays in SNC meteorites. *Icarus* 99: 28–41.
Gooding JL, Arvidson R, and Zolotov MYu (1992) Physical and chemical weathering. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 626–651. Tucson, AZ: University of Arizona Press.
Greeley R, Bender KC, Saunders RS, Schubert G, and Weitz CM (1997) Aeolian processes and features on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 547–590. Tucson, AZ: University of Arizona Press.
Greeley R and Iversen JD (1985) *Wind as a Geological Process on Earth, Mars, Venus and Titan*. Cambridge: Cambridge University Press, 333 pp.
Greeley R, Lankaster N, Lee S, and Thomas P (1992) Martian aeolian processes, sediments, and features. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 730–766. Tucson, AZ: University of Arizona Press.
Greeley R, Leach R, White B, Iversen J, and Pollack J (1980) Threshold windspeeds for sand on Mars: Wind tunnel simulations. *Geophysical Research Letters* 7: 121–124.
Greeley R, Marshall JR, and Pollack JB (1987) Physical and chemical modification of the surface of Venus by windblown particles. *Nature* 327: 313–331.
Grotzinger JP, Arvidson RE, Bell JF, et al. (2005) Stratigraphy and sedimentology of a dry to wet eolian depositional system, Burns formation, Meridiani Planum, Mars. *Earth and Planetary Science Letters* 240: 11–72.
Gupta SC, Ahrens TJ, and Yang W (2001) Shock-induced vaporization of anhydrite and global cooling from the K/T impact. *Earth and Planetary Science Letters* 188: 399–412.
Hashimoto GL and Abe Y (2001) Predictions of a simple cloud model for water vapor cloud albedo feedback on Venus. *Journal of Geophysical Research* 106: 14675–14690.
Hashimoto GL, Abe Y, and Sasaki S (1997) CO_2 amount on Venus constrained by a criterion of topographic-greenhouse instability. *Geophysical Research Letters* 24: 289–292.
Haskin LA, Wang A, Jolliff BL, et al. (2005) Water alteration of rocks and soils on Mars at the Spirit rover site in Gusev crater. *Nature* 436: 66–69.
Head JW and Wilson L (1986) Volcanic processes and landforms on Venus: Theory predictions and observations. *Journal of Geophysical Research* 91: 9407–9446.
Hecht MH, Kounaves SP, Quinn RC, et al. (2009) Detection of perchlorate and the soluble chemistry of Martian soil: Findings from the Phoenix Mars Lander. *Science* 325: 64–67.
Hoefen TM, Clark RN, Bandfield JL, et al. (2003) Discovery of olivine in the Nili Fossae region of Mars. *Science* 302: 627–630.
Huguenin RL (1973a) Photostimulated oxidation of magnetite 1. Kinetics and alteration phase identification. *Journal of Geophysical Research* 78: 8281–8493.
Huguenin RL (1973b) Photostimulated oxidation of magnetite 2. Mechanism. *Journal of Geophysical Research* 78: 8495–8506.
Huguenin RL (1974) The formation of goethite and hydrated clay minerals on Mars. *Journal of Geophysical Research* 79: 3895–3905.
Huguenin RL (1976) Mars: Chemical weathering as a massive volatile sink. *Icarus* 28: 203–212.
Huguenin RL (1982) Chemical weathering and the Viking biology experiments on Mars. *Journal of Geophysical Research* 87: 10069–10082.
Huguenin RL, Miller KJ, and Harwood WS (1979) Frost weathering on Mars: Experimental evidence for peroxide formation. *Journal of Molecular Evolution* 14: 103–132.
Jakosky BM and Haberle RM (1992) The seasonal behavior of water on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 969–1016. Tucson, AZ: University of Arizona Press.
James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, AZ: University of Arizona Press.
Johnson NM and Fegley B (2003) Tremolite decomposition on Venus II. Products, kinetics, and mechanism. *Icarus* 164: 317–333.
Johnson NM and Fegley B (2005) Phlogopite decomposition, water, and Venus. In: *36th Lunar and Planetary Science Conference (CD-OM)*, League City, TX: Lunar and Planetary Science, Abstract no. 1992.
Kasting JF (1991) CO_2 condensation and the climate of early Mars. *Icarus* 94: 1–13.
Kawaragi K, Sekine Y, Kadono T, et al. (2009) Direct measurements of chemical composition of shock-induced gases from calcite: An intense global warming after the Chicxulub impact due to the indirect greenhouse effect of carbon monoxide. *Earth and Planetary Science Letters* 282: 56–64.
Kemurdzhian AL, Brodskii PN, Gromov VV, et al. (1983) Preliminary results of determining the physical and mechanical properties of the soil of Venus by the Soviet automatic stations Venera 13 and Venera 14. *Cosmic Research* 21: 253–259.
Kieffer HH, Jakosky BM, and Snyder CW (1992) The planet Mars: From antiquity to the present. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1–33. Tucson, AZ: University of Arizona Press.
Kieffer HH and Zent AP (1992) Quasi-periodic climate change on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1180–1218. Tucson, AZ: University of Arizona Press.
Klein HP (1978) The Viking biological experiments on Mars. *Icarus* 34: 666–674.
Klingelhöfer G, Morris RV, Bernhardt B, et al. (2004) Jarosite and hematite at Meridiani Planum from Opportunity's Mössbauer spectrometer. *Science* 306: 1740–1745.
Klose KB, Wood JA, and Hashimoto A (1992) Mineral equilibria and the high radar reflectivity of Venus mountaintops. *Journal of Geophysical Research* 97: 16353–16369.
Knauth LP, Burt DM, and Wohletz KH (2005) Impact origin of sediments at the opportunity landing site on Mars. *Nature* 438: 1123–1128.
Krasnopolsky VA (2007) Chemical kinetic model for the lower atmosphere of Venus. *Icarus* 191: 25–37.
Krasnopolsky VA (2010) Spatially-resolved high-resolution spectroscopy of Venus 1. Variations of CO_2, CO, HF, and HCl at the cloud tops. *Icarus* 208: 539–547.
Krasnopolsky VA (2011) Atmospheric chemistry on Venus, Earth, and Mars: Main features and comparison. *Planetary and Space Science* 59: 952–964.
Krasnopolsky VA and Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294: 1914–1917.
Krasnopolsky VA, Maillard JP, and Owen TC (2004) Detection of methane in the Martian atmosphere: Evidence for life? *Icarus* 172: 537–547.
Krasnopolsky VA and Pollack JB (1994) H_2O–H_2SO_4 system in Venus' clouds and OCS, CO, and H_2SO_4 profiles in Venus' troposphere. *Icarus* 109: 58–78.
Ksanfomality LV, Gorshkova NV, and Khondyrev VK (1993) Wind velocity near the surface of Venus from acoustic measurements. *Cosmic Research* 21: 161–167.
Lammer F, Kasting JF, Chassefiére E, et al. (2008) Atmospheric escape and evolution of terrestrial planets and satellites. *Space Science Review* 139: 399–436.
Langevin Y, Poulet F, Bibring J-P, and Gondet B (2005) Sulfates in the North polar region of Mars detected by OMEGA/Mars express. *Science* 307: 1584–1586.
Leshin LA, Mahaffy PR, Webster CR, et al. (2013) Volatile, isotope, and organic analysis of Martian fines with the Mars Curiosity rover. *Science* 341: 1238937.
Lewis JS (1970) Venus: Atmospheric and lithospheric composition. *Earth and Planetary Science Letters* 10: 73–80.
Lewis JS and Prinn RG (1984) *Planets and Their Atmospheres*. Orlando, FL: Academic Press, 470 pp.
Marion GM, Catling DC, and Kargel JS (2009) Br/Cl partitioning in chloride minerals in the Burns formation on Mars. *Icarus* 200: 436–445.

McAdam AM, Franz BH, Sutter B, et al. (2014) Sulfur-bearing phases detected by evolved gas analysis of the Rocknest aeolian deposit, Gale Crater, Mars. *Journal of Geophysical Research: Planets* 119: 373–393.

McElroy MB, Kong TY, and Yung YL (1977) Photochemistry and evolution of Mars' atmosphere: A Viking perspective. *Journal of Geophysical Research* 82: 4379–4388.

McSween HY Jr. (2014) Mars. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 1. Oxford: Elsevier, 2nd edn.

McSween HY, McGlynn IO, and Rogers AD (2010) Determining the modal mineralogy of Martian soils. *Journal of Geophysical Research* 115: E00F12.

McSween HY Jr. and McLennan SM (2014) Mars. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 2, chapter 2.10. Oxford: Elsevier, 2nd edn.

McSween HY, Ruff SW, Morris RV, et al. (2006a) Alkaline volcanic rocks from the Columbia Hills, Gusev crater, Mars. *Journal of Geophysical Research* 111(E9): E09S91.

McSween HY, Ruff SW, Morris RV, et al. (2008) Mineralogy of volcanic rocks in Gusev crater, Mars: Reconciling Mössbauer, alpha particle X-ray spectrometer, and miniature thermal emission spectrometer spectra. *Journal of Geophysical Research* 113: E06S04.

McSween HY, Wyatt MB, Gellert R, et al. (2006b) Characterization and petrologic interpretation of olivine-rich basalts at Gusev crater, Mars. *Journal of Geophysical Research* 111: E02S10.

Mellon MT and Jakosky BM (1995) The distribution and behavior of Martian ground ice during past and present epochs. *Journal of Geophysical Research* 100: 11781–11799.

Michalski JR, Kraft MD, Sharp TG, Williams LB, and Christensen PR (2005) Mineralogical constraints on the high-silica Martian surface component observed by TES. *Icarus* 174: 161–177.

Milliken RE, Swayze GA, Arvidson RE, et al. (2008) Opaline silica in young deposits on Mars. *Geology* 36: 847–850.

Ming DW, Mittlefehldt DW, Morris RV, et al. (2006) Geochemical and mineralogical indicators for aqueous processes in the Columbia Hills of Gusev crater, Mars. *Journal of Geophysical Research* 111: E02S12.

Mitrofanov IG, Zuber MT, Litvak ML, et al. (2004) CO_2 snow depth and subsurface water–ice abundance in the northern hemisphere of Mars. *Science* 300: 2081–2084.

Moores JE and Schuerger AC (2012) UV degradation of accreted organics on Mars: IDP longevity, surface reservoir of organics, and relevance to the detection of methane in the atmosphere. *Journal of Geophysical Research* 117: E08008.

Morris RV, Klingelhöfer G, Schröder C, et al. (2006) Mössbauer mineralogy of rock, soil, and dust at Gusev crater, Mars: Spirit's journey through weakly altered olivine basalt on the plains and pervasively altered basalt in the Columbia Hills. *Journal of Geophysical Research* 111: E02S13.

Morris RV, Klingelhöfer G, Schröder C, et al. (2008) Iron mineralogy and aqueous alteration from Husband Hill through Home Plate at Gusev Crater, Mars: Results from the Mössbauer instrument on the Spirit Mars Exploration Rover. *Journal of Geophysical Research* 113: E12S42.

Morris RV and Lauer HV Jr. (1980) The case against UV photostimulated oxidation of magnetite. *Geophysical Research Letters* 7: 605–608.

Morris RV and Lauer HV Jr. (1981) Stability of goethite (–FeOOH) and lepidocrocite (–FeOOH) to dehydration by UV radiation – Implications for their occurrence on the Martian surface. *Journal of Geophysical Research* 86: 10893–10899.

Morris RV, Ruff RW, Gellert R, et al. (2010) Identification of carbonate-rich outcrops on Mars by the Spirit Rover. *Science* 329: 421–424.

Mueller RF (1964) A chemical model for the lower atmosphere of Venus. *Icarus* 3: 285–298.

Mueller RF (1968) Sources of HCl and HF in the atmosphere of Venus. *Nature* 220: 55–57.

Mukhin LM, Koscheev AP, Dikov YuP, Huth J, and Wänke H (1996) Experimental simulations of the photodecomposition of carbonates and sulphates on Mars. *Nature* 379: 141–143.

Murchie SL, Mustard JF, Ehlmann BL, et al. (2009) A synthesis of Martian aqueous mineralogy after 1 Mars year of observations from the Mars Reconnaissance Orbiter. *Journal of Geophysical Research* 114: E00D06.

Mustard JF, Poulet F, Gendrin A, et al. (2005) Olivine and pyroxene diversity in the crust of Mars. *Science* 307: 1594–1597.

Niles PB and Michalski J (2009) Meridiani planum sediments on Mars formed through weathering in massive ice deposits. *Nature Geoscience* 2: 215–220.

Nordstrom DK, Alpers CN, Ptacek CJ, and Blowes DW (2000) Negative pH and extremely acidic mine waters from Iron Mountain, California. *Environmental Science and Technology* 34: 254–258.

O'Connor JT (1968) Mineral stability at the Martian surface. *Journal of Geophysical Research* 73: 5301–5311.

Oppenheimer C (2003) Volcanic degassing. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 3, pp. 123–166. Oxford: Elsevier.

Osterloo MM, Anderson FS, Hamilton VE, and Hynek BM (2010) Geologic context of proposed chloride-bearing materials on Mars. *Journal of Geophysical Research* 115: E10012.

Owen T (1992) The composition and early history of the atmosphere of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 818–834. Tucson, AZ: University of Arizona Press.

Oyama VI and Berdahl BJ (1977) The Viking gas exchange experiment results from Chryse and Utopia surface samples. *Journal of Geophysical Research* 82: 4669–4676.

Oyama VI and Berdahl BJ (1979) A model of Martian surface chemistry. *Journal of Molecular Evolution* 14: 199–210.

Parkinson TD and Hunten DM (1972) Spectroscopy and aeronomy of O_2 on Mars. *Journal of the Atmospheric Sciences* 29: 1380–1390.

Pettengill GH, Campbell BA, Campbell DB, and Simpson RA (1997) Surface scattering and dialectical properties. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 527–546. Tucson, AZ: University of Arizona Press.

Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383.

Poch O, Noblet A, Stalport F, et al. (2013) Chemical evolution of organic molecules under Mars-like UV radiation conditions simulated in the laboratory with the "Mars organic molecule irradiation and evolution" (MOMIE) setup. *Planetary and Space Science* 85: 188–197.

Pope KO, Baines KH, Ocampo AC, and Ivanov BA (1997) Energy, volatile production, and climatic effects of the Chicxulub Cretaceous/Tertiary impact. *Journal of Geophysical Research* 102: 21645–21664.

Poulet F, Bibring J-P, Mustard JF, et al. (2005) Phyllosilicates on Mars and implications for early Martian climate. *Nature* 438: 623–627.

Prinn RG and Fegley B Jr. (1987) Bolide impacts, acid rain, and biospheric traumas at the Cretaceous–Tertiary boundary. *Earth and Planetary Science Letters* 83: 1–15.

Rieder R, Economou T, Wänke H, et al. (1997) The chemical composition of Martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* 278: 1771–1774.

Rieder R, Gellert R, Anderson RC, et al. (2004) Chemistry of rocks and soils at Meridiani Planum from the alpha particle X-ray spectrometer. *Science* 306: 746–1749.

Roach LR, Mustard JF, Lane MD, Bishop JL, and Murchie SL (2010) Diagenetic haematite and sulfate assemblages in Valles Marineris. *Icarus* 207: 659–674.

Ronov AB and Yaroshevsky AA (1976) A new model for the chemical structure of the Earth crust. *Geochemistry International* 13(6): 89–121.

Ruff SW, Farmer JD, Calvin WM, et al. (2011) Characteristics, distribution, origin, and significance of opaline silica observed by the Spirit rover in Gusev crater, Mars. *Journal of Geophysical Research* 116: E00F23.

Schaefer L and Fegley B Jr. (2004) Heavy metal frost on Venus. *Icarus* 168: 215–219.

Segura TL, McKay CP, and Toon OB (2012) An impact-induced, stable, runaway climate on Mars. *Icarus* 220: 144–148.

Segura TL, Toon OB, Colaprete A, and Zahnle K (2002) Environmental effects of large impacts on Mars. *Science* 298: 1977–1980.

Settle M (1979) Formation and deposition of volcanic sulfate aerosols. *Journal of Geophysical Research* 84: 8343–8354.

Shkuratov YuG, Kreslavsky MA, and Nikolayeva OV (1987) Albedo-color diagram of the Venusian surface and its interpretation. *Solar System Research* 21(2): 94–102.

Smith MD (2002) The annual cycle of water vapor on Mars as observed by the thermal emission spectrometer. *Journal of Geophysical Research* 84(E11): 5115.

Soderblom LA (1992) The composition and mineralogy of the Martian surface from spectroscopic observations. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 557–593. Tucson, AZ: University of Arizona Press.

Solomon SC, Bullock MA, and Grinspoon DH (1999) Climate change as a regulator of tectonics on Venus. *Science* 286: 87–90.

Squyres SW, Arvidson RE, Bell JF III, et al. (2012) Ancient impact and aqueous processes at Endeavour crater, Mars. *Science* 336: 570–576.

Squyres SW, Knoll AH, Arvidson RE, et al. (2006) Two years at Meridiani Planum: Results from the opportunity rover. *Science* 313: 1403–1407.

Steiger M, Linnow K, Ehrhard D, and Rohde M (2011) Decomposition reactions of magnesium sulfate hydrates and phase equilibria in the $MgSO_4$–H_2O and Na^+–Mg^{2+}–Cl^-–SO_4^{2-}–H_2O systems with implications for Mars. *Geochimica et Cosmochimica Acta* 75: 3600–3626.

Surkov YuA, Barsukov VL, Moskalyeva LP, Kharyukova VP, and Kemurdzhian AL (1984) New data on the composition, structure, and properties of Venus rock obtained by Venera-13 and Venera-14. *Journal of Geophysical Research: Proceedings of the 15th LPSC* 89: 393–402.

Surkov YuA, Kirnozov FF, Glazov VN, Dunchenko AG, Tatsy LP, and Sobornov OP (1987) Uranium, thorium, and potassium in the Venusian rocks at the landing sites of Vega-1 and Vega-2. *Journal of Geophysical Research: Proceedings of the 17th LPSC* 92: 537–540.

Surkov YuA, Moskalyova LP, Kharyukova VP, Dudin AD, Smirnov GG, and Zaitseva SYe (1986) Venus rock composition at the Vega-2 landing site. *Journal of Geophysical Research: Proceedings of the 16th LPSC* 91: 215–218.

Symonds RB, Rose WI, Bouth GJS, and Gerlach TM (1994) Volcanic-gas studies: Methods, results, and applications. In: Carroll MR and Holloway JR (eds.) *Volatiles in Magmas*, pp. 1–66. Washington, DC: Mineralogical Society of America.

Tanner SB, Kerrick DM, and Lasaga AC (1985) Experimental kinetic study of the reaction: Calcite + quartz → wollastonite + carbon dioxide, from 1 to 3 kilobars and 500° to 850 °C. *American Journal of Science* 285: 577–620.

Terradellas J and Bonnetain L (1973) Nature of the chemical reactions during the reaction of sulfur dioxide with calcium carbonate. *Bulletin de la Société Chimique de France* 6: 1903–1908.

Thomas P, Squyres S, Herkenhoff K, Howard A, and Murray B (1992) Polar deposits on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 767–798. Tucson, AZ: University of Arizona Press.

Tian F, Claire MW, Haqq-Misra JD, et al. (2010) Photochemical and climate consequences of sulfur outgassing on early Mars. *Earth and Planetary Science Letters* 295: 412–418.

Toon OB, Zahnle K, Morrison D, Turco RP, and Covey C (1997) Environmental perturbations caused by the impacts of asteroids and comets. *Reviews of Geophysics* 35: 41–78.

Treiman AH and Bullock MA (2011) Mineral reaction buffering of Venus' atmosphere: A thermochemical constraint and implications for Venus-like planets. *Icarus* 217: 534–541.

Urey HC (1952) *The Planets*. New Haven: Yale University Press.

Vaniman DT, Bish DL, Ming DW, et al. (2014) Mineralogy of a mudstone at Yellowknife Bay, Gale crater, Mars. *Science* 343: 1243480. http://dx.doi.org/10.1126/science.1243480.

Volkov VP (1992) Volatiles in atmosphere and crust. In: Barsukov VL, Basilevsky AT, Volkov VP, and Zharkov VN (eds.) *Venus Geology, Geochemistry, and Geophysics*, pp. 200–210. Tucson, AZ: University of Arizona Press.

Volkov VP, Zolotov MYu, and Khodakovsky IL (1986) Lithospheric–atmospheric interaction on Venus. In: Saxena SK (ed.) *Chemistry and Physics of Terrestrial Planets*, pp. 136–190. New York: Springer.

Von Zahn U, Kumar S, Niemann H, and Prinn R (1983) Composition of the Venus atmosphere. In: Hunten DM, Colin L, Donahue TM, and Moroz VM (eds.) *Venus*, pp. 299–430. Tucson, AZ: University of Arizona Press.

Wang A, Freeman JJ, Chou I-M, and Jolliff BL (2011) Stability of Mg-sulfates at −10 °C and the rates of dehydration/rehydration processes under conditions relevant to Mars. *Journal of Geophysical Research* 116: E12006.

Wänke H and Dreibus G (1987) Volatiles on Earth and Mars: A comparison. *Icarus* 71: 225–240.

Webster CR, Mahaffy PR, Atreya SK, et al. (2013) Low upper limit to methane abundance on Mars. *Science* 342: 355–357.

Wood JA (1997) Rock weathering on the surface of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 637–666. Tucson, AZ: University of Arizona Press.

Yen AS, Gellert R, Schröder C, et al. (2005) An integrated view of the chemistry and mineralogy of Martian soils. *Nature* 436: 49–54.

Yen AS, Murray B, Rossman GR, and Grunthaner FJ (1999) Stability of hydroxylated minerals on Mars: A study on the effects of exposure to ultraviolet radiation. *Journal of Geophysical Research* 104: 27031–27041.

Yung YL and DeMore WD (1999) *Photochemistry of Planetary Atmospheres*. New York: Oxford University Press, 456 pp.

Zent AP and McKay CP (1994) The chemical reactivity of the Martian soil and implications for future missions. *Icarus* 108: 146–157.

Zolotov MYu (1996) A model for the thermal equilibrium of the surface Venusian atmosphere. *Geochemistry International* 33(10): 80–100.

Zolotov MYu, Fegley B Jr., and Lodders K (1997) Hydrous silicates and water on Venus. *Icarus* 130: 475–494.

Zolotov MYu and Mironenko MV (2007) Timing of acid weathering on Mars: A kinetic-thermodynamic assessment. *Journal of Geophysical Research* 112: E07006. http://dx.doi.org/10.1029/2006JE002882.

Zolotov MYu and Volkov VP (1992) Chemical processes on the planetary surface. In: Barsukov VL, Basilevsky AT, Volkov VP, and Zharkov VN (eds.) *Venus Geology, Geochemistry, and Geophysics*, pp. 177–200. Tucson, AZ: University of Arizona Press.

This page intentionally left blank

10.13 Planetary Atmospheres

DC Catling, University of Washington, Seattle, WA, USA

© 2015 Elsevier B.V. All rights reserved.

10.13.1	**Introduction and Scope**	429
10.13.2	**Composition and Vertical Structure**	430
10.13.2.1	The Observed Compositional Diversity of Atmospheres	430
10.13.2.2	Nomenclature for the Vertical Structure of Atmospheres	432
10.13.2.3	The Equation of State, Atmospheric Composition, and the Hydrostatic Equation	434
10.13.2.4	Convection and Stability	435
10.13.2.5	Condensables and the Clausius–Clapeyron Equation	437
10.13.3	**Radiation**	438
10.13.3.1	Planetary Energy Balance, Blackbody Radiation, and the Greenhouse Effect	438
10.13.3.2	Radiative Transfer: Shortwave and Infrared	441
10.13.3.3	Optical Thickness and Thinness	444
10.13.3.4	Radiative–Convective Equilibrium	445
10.13.3.5	Absorption and Emission by Atmospheric Gases	450
10.13.3.5.1	Rotational transitions	451
10.13.3.5.2	Vibrational transitions	451
10.13.3.5.3	Collision-induced absorption	451
10.13.3.5.4	Broadening	451
10.13.3.5.5	Continuum	452
10.13.3.5.6	Numerical radiative transfer	452
10.13.4	**Photochemistry**	452
10.13.4.1	Basic Principles of Chemical Kinetics	452
10.13.4.2	Why Free Radicals Are So Important	453
10.13.4.3	Photolysis	453
10.13.4.4	Earth's Atmospheric Composition and Chemistry	454
10.13.4.5	CO_2 Atmospheres: Venus and Mars	455
10.13.4.6	Reducing Atmospheres: Giant Planets and Titan	455
10.13.5	**Atmospheric Dynamics**	455
10.13.5.1	Equations of Motion	455
10.13.5.2	Balance Regimes: Geostrophic, Cyclostrophic, and Thermodynamic	457
10.13.5.2.1	Geostrophic balance	457
10.13.5.2.2	Cyclostrophic balance	457
10.13.5.2.3	Thermodynamic balance	458
10.13.5.3	Zonal Mean Meridional Circulation: Hadley Cells and Thermally Driven Jets	458
10.13.5.4	Eddy-Driven Jet Streams	461
10.13.5.5	Planetary Waves and Barotropic and Baroclinic Instabilities	462
10.13.5.6	Deep and Shallow Atmospheres on Giant Planets	464
10.13.5.7	Buoyancy Waves and Thermal Tides	465
10.13.5.8	Superrotation	465
10.13.6	**Outstanding Questions**	466
Acknowledgments		467
References		467

10.13.1 Introduction and Scope

The enormous topic of planetary atmospheres ranges from the origin of atmospheres to the fluid dynamics of small weather systems, so here, I restrict the scope to broad essentials at a level suitable for a PhD research student. I assume knowledge of calculus, plus basic physics and chemistry, but little else. The audience could also include researchers in allied fields – for example, planetary geologists, geochemists, or astronomers. I focus on atmospheres in the solar system. Exoplanets represent a new frontier that should reveal a larger diversity of atmospheres (Seager, 2010; Seager and Deming, 2010; Seager et al., 2010). But at the moment, only a little can be said about exoplanet atmospheres with any certainty. Currently, observational data from such atmospheres are arguably much sparser than data were for atmospheres in our solar system before the Space Age. Rather than speculate, I concentrate on key principles associated with the planetary atmospheres that we know best. In fact, these are what will be needed to interpret exoplanet atmospheres correctly (see **Chapter 10.21**).

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00185-8

Until the last 5 years or so, introductory syntheses about atmospheres were mostly confined to the Earth (e.g., Houghton, 2002; Wallace and Hobbs, 2006), with a few exceptions (Chamberlain and Hunten, 1987; Yung and DeMore, 1999), but now, the coverage of planetary atmospheres is more available. A primer by Ingersoll (2013) is a good read for all levels, while Taylor (2010) provides a useful introduction aimed at mid-level undergraduates. At the research level, general texts are Pierrehumbert (2010), Sanchez-Lavega (2010), and Catling and Kasting (in press). In a more specialized area, Yung and Demore (1999) provided a detailed monograph on photochemistry. Recent papers have also reviewed the probable atmospheric circulation regimes of exoplanets (Showman et al., 2010, 2013).

The structure of this review is as follows. In order to discuss atmospheres, I start with an overview of composition and basic physical properties, including the layers of atmospheres. Thus, **Section 10.13.2** describes structure and thermodynamics. Atmospheric composition determines physical properties, which in turn affects climate through the greenhouse effect. A planet's climate requires a grasp of the radiative transfer of energy, which is discussed in **Section 10.13.3**. A brief overview of the photochemistry of atmospheres is then given in **Section 10.13.4** to provide an appreciation of the globally averaged bulk composition and vertical distribution of constituents. However, atmospheric motions transport heat and constituents, including condensable species, and are indispensable for several aspects of climate and chemistry. Principles of atmospheric motion are described in **Section 10.13.5**. Together, these various topics – thermodynamics, radiation, chemistry, and dynamics – provide the essentials of planetary atmospheres. Finally, in **Section 10.13.6**, I very briefly note some key outstanding questions.

One topic is missing: atmospheric origin and evolution. This subject is of profound interest but involves considerable geochemistry and familiarity with all the topics of this review, so the reader is referred to Catling and Kasting (in press) that comprehensively reviews this area (see **Chapter 10.14**).

10.13.2 Composition and Vertical Structure

10.13.2.1 The Observed Compositional Diversity of Atmospheres

A physically meaningful classification of the composition of planetary atmospheres falls into two chemical categories: reducing and oxidizing. A redox (reduction–oxidation) state indicates the chemical character of atmospheres, for example, whether elements such as carbon are likely to be found in their reduced form (methane, CH_4) or oxidized variety (carbon dioxide, CO_2).

The solar system contains several examples of oxidizing and reducing atmospheres. Reducing atmospheres are those that are relatively rich in reducing gases, typically hydrogen-bearing gases, such as hydrogen itself (H_2), methane (CH_4), and possibly ammonia (NH_3). Examples include all the giant planets and Saturn's largest moon, Titan (**Table 1**). In principle, an atmosphere dominated by carbon monoxide (CO) would also be reducing (perhaps applicable to some hot giant exoplanet atmospheres (Moses et al., 2011) or early Mars (Zahnle et al., 2008)), although we do not have an example in the modern solar system, and such an atmosphere would likely contain considerable H_2 as well. The atmosphere of the early Earth was also reducing, albeit weakly, before its oxygenation at 2.4 Ga (Ga = giga anna = billions of years ago) (Catling, 2014; Lyons et al., 2014). In contrast, oxidizing atmospheres are relatively poor in reducing hydrogen-bearing gases and include modern Earth and Mars (**Table 2**). Molecular O_2 has a volume abundance of 21 and 0.15% on Earth and Mars, respectively, which greatly exceeds the amount of hydrogen-bearing reducing gases of 1.8 ppmv (parts per million by volume) CH_4 on Earth (Hartmann et al., 2013) and 15 ± 5 ppmv H_2 on Mars (Krasnopolsky and Feldman, 2001). Venus is unusual. The upper atmosphere of Venus is oxidizing, with a tenuous ozone (O_3) layer at 90–105 km altitude on the nightside (Montmessin et al., 2011), while the atmosphere below ~20 km altitude is arguably weakly reducing because sulfur is converted into various reduced forms (Krasnopolsky and Pollack, 1994). Noting sulfur oxidation numbers in square brackets, these sulfur species include sulfur dioxide (SO_2) [+4], carbonyl sulfide (OCS)[−2], hydrogen sulfide (H_2S) [−2], and elemental sulfur allotropes of 1–8 atoms ($S–S_8$)[0]. The OCS is sourced from weathering of sulfides on the surface, or volcanic outgassing, or the reduction of sulfur by carbon monoxide (e.g., $S_2 + CO$), which is present at ~20 ppmv below 20 km (Fegley, 2014). Unlike Earth and Mars, O_2 is absent in Venus' atmosphere with an upper limit of <0.3 ppmv.

Broadly speaking, planetary formation theories from a Sun-like nebula (reviewed by Chambers, 2014) favor more hydrogen-rich atmospheres far from a star and more oxidizing atmospheres nearby, but this can be confounded in an extrasolar system if a planet's orbital distance changes in planetary migration. Also, the very early rocky planets in our own solar system may have had hydrogen-rich atmospheres (Abe, 2011).

Redox also indicates the possible chemistry of aerosols, which are fine particles found in atmospheres. In general, oxidizing atmospheres with sulfurous gases tend to produce sulfate aerosols at altitude (McGouldrick et al., 2011), such as in the Earth's stratosphere (Crutzen, 1976; Robock, 2000), on Venus (Thompson et al., 2011), or on Mars when it was volcanically active (Settle, 1979; Smith et al., 2014; Tian et al., 2010). In contrast, reducing atmospheres generate hydrocarbon aerosols, including the giant planets (Irwin, 2009; Moses et al., 2004; West et al., 1991, 2004, 2009), Titan (Krasnopolsky, 2010; Lavvas et al., 2008a,b; Wilson and Atreya, 2004), and probably the early Earth (Pavlov et al., 2001; Wolf and Toon, 2010; Zerkle et al., 2012).

In addition to the atmospheres discussed in the preceding text, some smaller bodies possess very thin atmospheres. Exospheres exist directly above the surfaces of several moons, which are defined as atmospheric layers of such low density that they are virtually collisionless. We find the following thin atmospheres, where the surface pressure is given in bar (1 bar = 10^5 Pa) in parentheses:

(i) N_2-rich atmospheres above surfaces covered in N_2 ice: Triton ($\sim(14–20) \times 10^{-6}$), the largest moon of Neptune, and Pluto ($\sim 15 \times 10^{-6}$)

Table 1 The composition of reducing atmospheres

Species	*Jupiter*	*Saturn*	*Uranus*	*Neptune*	*Titan*
H_2	$86.4 \pm 2.6\%$	$88 \pm 2\%$	$82.5 \pm 3.3\%$	$80 \pm 3.2\%$	0.099%
He	$13.6 \pm 0.3\%$	$12 \pm 2\%$	$15.2 \pm 3.3\%$	$19.0 \pm 3.2\%$	
CH_4[a]	$0.204 \pm 0.049\%$	$0.47 \pm 0.2\%$	~2.3% (below CH_4 clouds)	~1–2% (below CH_4 clouds)	5.65% (1.48% in stratosphere)
$^{13}CH_4$	$1.9 \pm 0.1 \times 10^{-5}$	$5.1 \pm 0.2 \times 10^{-5}$			0.062%
N_2	–	–	–	–	94.2%
NH_3[a]	$5.74 \pm 0.22 \times 10^{-4}$ (8.9–11.7 bar, below NH_3 clouds)	$\sim 1.1 \times 10^{-4}$ ($p > 1$ bar, below NH_3 clouds)	$< 1 \times 10^{-7}$	6×10^{-7} (at 6 bar)	
H_2O[a]	$4.23 \pm 1.38 \times 10^{-4}$ (at 17.6–20.9 bar)	2.3×10^{-7} (upper troposphere)	$6–14 \times 10^{-9}$ (at $p < 0.03$ mbar)	$1.7–4.1 \times 10^{-9}$ (at $p < 0.6$ mbar)	$< 10^{-8}$
H_2S[a]	$6.7 \pm 0.4 \times 10^{-5}$ (at $p > 16$ bar)	$< 0.4 \times 10^{-6}$	$< 8 \times 10^{-7}$	$< 3 \times 10^{-6}$	–
Ne	1.99×10^{-5}				$< 2 \times 10^{-5}$ ^{20}Ne
Ar	$1.57 \pm 0.35 \times 10^{-5}$				3.4×10^{-5} ^{40}Ar 2.1×10^{-7} ^{36}Ar
C_2H_6	$5.8 \pm 1.5 \times 10^{-6}$ (stratosphere)	$7.0 \pm 1.5 \times 10^{-6}$ (stratosphere)	1.5×10^{-5} (at 0.5–1 mbar)	1.5×10^{-6} (stratosphere)	$\leq 1 \times 10^{-5}$
PH_3	$1.04 \pm 0.1 \times 10^{-6}$	$5.9 \pm 0.2 \times 10^{-6}$ ($p > 0.5$ bar)	$< 8.3 \times 10^{-7}$		-
C_2H_2	$\sim 4 \times 10^{-8}$ (at 5 mbar)	$3.0 \pm 0.2 \times 10^{-7}$ (at 1 mbar)	$\sim 1.0 \times 10^{-8}$ (stratosphere)	$6.0^{+14.0}_{-4.0} \times 10^{-9}$ (stratosphere)	$\leq 2 \times 10^{-6}$
C_2H_2	3.5×10^{-10} (at $p < 10$ mbar)	3.4×10^{-10} (at $p < 10$ mbar)	$4.0 \pm 0.5 \times 10^{-11}$ (at 0.1 mbar)	6×10^{-8} (at $p < 5$ mbar)	10^{-8} (above stratospheric condensation)
CO	$1.6 \pm 0.3 \times 10^{-9}$ (at 5–8 bar)	$< 10^{-7}$ (stratosphere)	$< 4 \times 10^{-8}$ (stratosphere)	$6.5 \pm 3.5 \times 10^{-6}$ (stratosphere)	4.7×10^{-5}
D/H	$2.3–2.6 \times 10^{-5}$	$1.7^{+0.75}_{-0.45} \times 10^{-5}$			$1.35 \pm 0.3 \times 10^{-4}$ in H_2 $1.32^{+0.15}_{-0.11} \times 10^{-4}$ in CH_4
$^{13}C/^{12}C$	0.0108 ± 0.0005	0.011			0.01099 ± 0.00017 in CH_4

[a]Condensable species, so levels are indicated.

Gases are given as volume mixing ratios. CH_4 is condensable on Titan, Uranus, and Neptune. H_2O and NH_3 are condensable on the giant planets, while H_2S reacts with NH_3 to form ammonium hydrosulfide (NH_4SH) clouds.

Source: Lodders K (2010) Atmospheric chemistry of the gas giant planets. *Geochemical News* 142: 1–11; Atreya SK, Mahaffy PR, Niemann HB, Wong MH, and Owen TC (2003) Composition and origin of the atmosphere of Jupiter – An update, and implications for the extrasolar giant planets. *Planetary and Space Science* 51: 105–112; Irwin P (2009) *Giant Planets of Our Solar System: Atmospheres, Composition, and Structure*, 2nd edn. Chichester, UK: Springer; Niemann HB, Atreya SK, Demick JE, Gautier D, Haberman JA, Harpold DN, Kasprzak WT, Lunine JI, et al. (2010) Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini–Huygens probe gas chromatograph mass spectrometer experiment. *Journal of Geophysical Research* 115.

(ii) O_2-rich exospheres above moons covered in water ice: Jupiter's moons, Europa ($\sim 10^{12}$–10^{13} O_2), (Smyth and Marconi, 2006), Ganymede ($\sim 10^{-12}$ O_2) (Hall et al., 1998), and Callisto ($\sim 10^{-12}$ O_2 possibly + CO_2) (Carlson, 1999; Kliore et al., 2002; Liang et al., 2005), and Saturn's moons, Rhea ($\sim 10^{-12}$ O_2 and CO_2) (Teolis et al., 2010) and Dione ($\sim 10^{-12}$ O_2) (Tokar et al., 2012)

(iii) Volcanogenic atmospheres: SO_2 on Io ($\sim 10^{-7}$–10^{-9}), a moon of Jupiter, and a water vapor–CO_2 exosphere on Enceladus, a moon of Saturn

The tenuous N_2-rich atmospheres (with minor CH_4) on Triton and Pluto arise from vapor equilibrium with N_2 (and CH_4) ice at prevailing surface temperatures. The O_2-rich exospheres on the moons of Jupiter and Saturn come from the breakdown of water ice on the surfaces of these moons by radiolysis and sputtering (Johnson et al., 2009). Radiolysis is the chemical alteration of ices by charged particles, and sputtering is when charged particles collide and eject atoms or molecules. The faster escape of the hydrogen compared with that of the oxygen leaves oxygen lingering behind in an exosphere. The charged particles come from magnetospheres, which are envelopes of ions around planets controlled by the planetary magnetic fields.

Volcanic small bodies have a different class of atmosphere. The very thin SO_2 atmosphere on Io, a rocky body, is sourced from high-temperature volcanism at a rate of 10^5 kg s^{-1} SO_2 (Lellouch et al., 2003). The SO_2 freezes to form frost away from the hot vents (Schmitt et al., 1994). Elemental sulfur is also volcanically outgassed, with $SO_2/S_2 \sim 3$–12 in the plume of the volcano Pele (Spencer et al., 2000). In contrast to Io, Enceladus is covered in water ice and affected by cryovolcanism, that is, a form of volcanism where ice turns to vapor or gets ejected as a result subsurface heating, unlike volcanism that involves molten rock. Plumes at the south pole of Enceladus release icy particles at a rate of 200 ± 30 kg s^{-1} (Hansen et al., 2011) from parallel lineaments called 'tiger stripes' (~500 m deep, ~130 km long, 2 km wide) (Porco et al., 2006), causing a very asymmetrical exosphere (Tenishev et al., 2010).

The very thin atmospheres described in the preceding text are mentioned for completeness. Henceforth, I restrict most of the discussion to thicker atmospheres, the thinnest being that of Mars.

Table 2 The composition and properties of oxidizing atmospheres (noting that the lower atmosphere of Venus could be considered somewhat reducing)

Parameter	*Venus*	*Earth*	*Mars*
Mean surface pressure (bar)	95.6	1.0	0.006
Mean surface temperature (K)	735	288	218
Mass relative to the Earth (5.97×10^{24} kg)	0.815	1.0	$0.107 \approx 1/9$
Mean radius relative to the Earth (6371 km)	0.950	1.0	$0.532 \approx 1/2$
Key gases in atmosphere (by volume)	CO_2 96.5% N_2 3.5% SO_2^c ● 150 ± 30 ppm (22–42 km) ● 25 ± 150 ppm (12–22 km) ^{40}Ar 70 ± 25 ppm $^{36+38}Ar$ 75 ± 35 ppm H_2O^c 30 ± 15 ppm (5–45 km) CO^c ● 45 ± 10 ppm (cloud top) ● 17 ± 1 ppm (12 km) He 12 (+24/−8) ppm Ne 7 ± 3 ppm H_2^c 2.5 ± 1 ppm (50–60 km) HCl 0.4 ppm (70 km) ^{84}Kr 50 ± 25 ppb	N_2[a] 78.084% O_2[a] 20.946% H_2O 0.1 ppm – 4% (varies) Ar 9340 ppm CO_2[a,b] ● ~280 ppm (pre-industrial) ● 396.5 ppm (year 2013) Ne 18.18 ppm ^4He 5.24 ppm CH_4[a] 1.7 ppm Kr 1.14 ppm H_2[a] 0.55 ppm N_2O[a] ~320 ppb CO[a] 125 ppb	CO_2 96.0 ± 0.7% N_2 1.89 ± 0.01% Ar 1.93 ± 0.01% O_2 0.145 ± 0.09% CO 0.06% H_2O ~0.03% (varies) He 10 ppm H_2 15 ± 5 ppm Ne 2.5 ppm Kr 0.3 ppm O_3 0–80 ppb H_2O_2 0–40 ppb SO_2[d] <0.3 ppb

[a]Biologically cycled and influenced.
[b]Currently increasing about 2 ppm per year because of anthropogenic emissions.
[c]Altitude-dependent.
[d]Indicates lack of volcanic outgassing.
Source: Lodders K and Fegley B (1998) *The Planetary Scientist's Companion*. New York: Oxford University Press; Pollack JB (1991) Kuiper prize lecture: Present and past climates of the terrestrial planets. *Icarus* 91: 173–198; Krasnopolsky VA and Lefevre F (2013) Chemistry of the atmospheres of Mars, Venus, and Titan. In: Mackwell SJ, Simon-Miller AA, Harder JW, and Bullock MA (eds.) *Comparative Climatology of Terrestrial Planets*, pp. 231–275. Tucson, AZ: University of Arizona Press; Mahaffy PR, Webster CR, Atreya SK, et al. (2013) Abundance and isotopic composition of gases in the martian atmosphere from the Curiosity Rover. *Science* 341: 263–266.

10.13.2.2 Nomenclature for the Vertical Structure of Atmospheres

Along with composition, a basic description of any atmospheres involves atmospheric structure. The term conventionally refers to the vertical temperature profile, which provides information about how chemical and physical processes change with altitude. The average thermal profile is also the basis for naming atmospheric layers using a system developed for Earth in the early 1900s (Lindemann and Dobson, 1923; Martyn and Pulley, 1936); thus, the nomenclature works with only some degree of success for other bodies. New atmospheric layers begin where there is a temperature inflection (**Figure 1**). On Earth, the five layers are the following: (i) the troposphere (tropos – Greek for the 'turning' of convection), where temperature declines from the surface to the tropopause (pause – Greek for 'stop'), which lies at an average pressure of 0.16 bar (Sausen and Santer, 2003). The tropopause altitude is 0.2 bar (11 km) in the US Standard Atmosphere but varies in detail with latitude from 0.1 to 0.3 bar (tropics to pole) (Hoinka, 1998). (ii) The stratosphere (stratus–Greek for 'layered'), where temperature increases from the tropopause to the stratopause at ~50 km altitude where the air pressure is ~100 Pa (1 mbar). (iii) The mesosphere from the stratopause to the mesopause, at ~85 km altitude, where the air pressure is about 1–0.1 Pa (0.01–0.001 mbar). (iv) The thermosphere from the mesopause to the thermopause (or exobase) at that varies over altitudes of ~250–500 km as solar ultraviolet (UV) flux changes over cycles ranging from daily to the 11-year solar cycle. (v) The exosphere, which blends into interplanetary space.

Convection dominates lower tropospheres. Solar radiation heats the surface of rocky planets or is absorbed in the deeper portions of tropospheres on giant planets. Consequently, warm buoyant air rises to altitudes of lower pressure where it expands and cools. Cool air parcels sink and warm. The net effect of such convection is a decline of temperature with altitude. On Earth, the rate that temperature decreases, the lapse rate, averages ~6 K km^{-1} globally. In the upper troposphere where the air is thinner, radiation transfers energy instead of convection above a radiative–convective boundary. Higher still, emission of thermal infrared (IR) radiation from the air to space becomes so efficient that a temperature minimum or tropopause is achieved. The tropopause occurs near an average pressure of roughly 0.1 bar on Earth, Titan, and the giant planets. The similarity is because of a common scaling of broadband IR opacity and pressure. In upper tropospheres, the IR opacity is pressure-dependent in the same way – varying approximately with the square of the pressure – while all these atmospheres are roughly similar in their strong opaqueness to IR at deeper pressures, such as 1 bar,

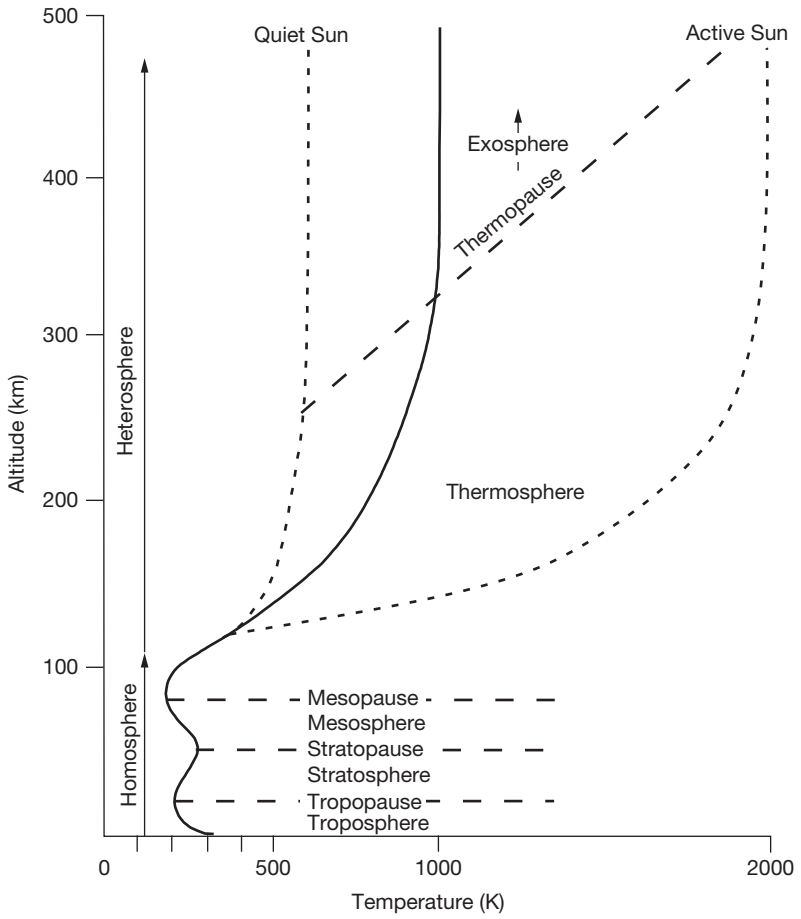


Figure 1 The terminology for vertical layers of the Earth's atmosphere, shown schematically. (Drawn by the author.)

despite the big differences in atmospheric composition and gravity (Robinson and Catling, 2014) (see **Section 10.13.3.4**).

In stratospheres, the absorption of shortwave solar radiation causes a temperature inversion, defined as a temperature increase with altitude. On Earth, the ozone absorbs UV radiation, whereas on the giant planets and Titan, aerosol particles and methane absorb at wavelengths in the UV, visible, and near-IR. Mars and Venus, which have predominantly CO_2 atmospheres, do not have strong shortwave absorbers above their tropospheres, so in the global average, they lack stratospheric inversions and tropopause minima (see **Section 10.13.3.4**).

Above the Earth's stratosphere, CO_2 emission to space causes the mesosphere to have the lowest atmospheric temperatures. A cold mesopause forms the lower boundary of the thermosphere where temperature increases with altitude. Mesopauses occur around ~1–0.1 Pa in planetary atmospheres in the solar system. They are cold because the air is so thin in upper mesospheres that energy-exchanging collisions between molecules become less frequent than photon emission at a level of radiative relaxation (Curtis and Goody, 1956; López-Puertas and Taylor, 2001). Consequently, excited molecules lose energy by radiation rather than by collisions.

In thermospheres, absorption of extreme ultraviolet radiation (EUV) (with wavelength shorter than ~120 nm) causes temperature to increase. Molecules are broken apart by EUV and photochemical reactions, and EUV and x-rays ionize both molecules and atoms. In the tenuous gas, infrequent collisions between electrons and ions create an ionosphere, a shell of electrons and ions that extends on Earth from the mesosphere into the exosphere. The Earth's hot thermosphere (~1000 K at 250 km altitude) reflects the slow rate of removing heat. Atoms and atomic ions do not have vibrational or rotational transitions and so are poor radiators, unlike molecules. Instead, heat is transferred inefficiently by conduction downward to the cold mesopause where radiative cooling is effective. Unlike the Earth, the upper thermospheres on Mars and Venus are relatively cold at ~270 K for Mars and ~250 K for Venus, because their ~96% CO_2 causes efficient radiative cooling to space (**Section 10.13.4.5**). Finally, exospheres lie above thermospheres and are where atmospheres become collisionless.

Chemically, we can divide atmospheres into two big layers: a homosphere where bulk gases are well mixed up to a homopause and a heterosphere above the homopause where molecular diffusion becomes more important than turbulent mixing. The homopause occurs where the diffusion separation time and turbulent mixing time of gases are equal. This tends to occur within a few orders of magnitude of 0.01 Pa (or 0.1 μbar). On Earth, the homopause is at ~100 km height. Above, at 250–1000 km altitude, diffusive separation causes atomic oxygen (O) to become the dominant species, while higher up, helium dominates, and then, hydrogen above 2500 km height.

Species escape from the base of exospheres (the exobase) if their upward velocity exceeds the escape velocity v_{esc}, which is given by a balance between gravitational energy and kinetic energy (KE):

$$\frac{1}{2}mv_{esc}^2 = \frac{GMm}{r} \Rightarrow v_{esc} = \sqrt{\frac{2GM}{r}} \Rightarrow \underbrace{v_{esc} \propto \sqrt{\frac{a^3}{a}} \propto a}_{\text{if same density body of radius } a} \quad [1]$$

Here, r is the radial distance, G is the universal gravitational constant (6.672×10^{-11} N m^2 kg^{-2}), M is the mass of the planet, a is the planet's radius, and m is the molecular mass. Today, hydrogen escapes efficiently from Earth at a small rate ~3 kg s^{-1}. Hydrogen atoms at the exobase in the tail of the Maxwell–Boltzmann velocity distribution have enough thermal velocity to evaporate and escape even though the root-mean-square speed (given by $1/2mv^2 = (3/2)kT$, where $k = 1.381 \times 10^{-23}$ J K^{-1} is Boltzmann's constant and T is exobase temperature) is less than the Earth's escape velocity. Alternatively, the hydrogen is lost when chemical or charged particle reactions impart speed to hydrogen atoms. The rate-limiting step is not a loss from the exobase itself but diffusion from the homopause to the exobase. Consequently, hydrogen is said to be lost at an upper limit by diffusion-limited escape (Hunten, 1973; Walker, 1977). Diffusion-limited escape of hydrogen is an excellent theory given that data suggest that it happens on Venus, Earth, Mars, and Titan.

For completeness, there are three ways that gases escape from planets: (i) thermal escape, with end-member approximations of thermal evaporation (as mentioned in the preceding text) called Jean's escape at one end and pressure-driven hydrodynamic escape at the other (e.g., Catling and Zahnle, 2009; Johnson et al., 2013; Walker, 1982), (ii) nonthermal escape where an atom or molecule is boosted to escape velocity because of a chemical reaction or ionic interaction, and (iii) impact erosion where atmospheric gases are expelled as a result of the large body impacts. Helium escapes from the Earth today through nonthermal escape. However, bodies can lose their atmospheres in bulk. Theory demonstrates that the two mechanisms that most efficiently strip away substantial atmospheres are hydrodynamic escape driven by stellar irradiation (Lammer et al., 2008; Sekiya et al., 1980, 1981; Watson et al., 1981; Zahnle and Kasting, 1986; Zahnle et al., 1990) and impact erosion (Griffith and Zahnle, 1995; Melosh and Vickery, 1989; Walker, 1986; Zahnle et al., 1992). Currently, in the solar system, only Pluto's atmosphere is believed to be undergoing hydrodynamic escape (Trafton et al., 1997). The reader is referred to Catling and Kasting (in press) and Lammer (2013) for more discussion of escape processes.

10.13.2.3 The Equation of State, Atmospheric Composition, and the Hydrostatic Equation

The most basic physics of atmospheres involves connecting the distributions of pressure, temperature, and density with (i) an equation of state and (ii) a hydrostatic equation.

Except for deep within giant-planet atmospheres, the ideal gas law works well as for (i)

$$p = nkT = \underbrace{(n\overline{m})}_{\rho}\left(\frac{k}{\overline{m}}\right)T \Rightarrow p = \rho\overline{R}T \quad [2]$$

Here, k is Boltzmann's constant, $\overline{m}$ is the mean molecular mass, n is the number density or number of molecules per unit volume, and $\rho = n\overline{m}$ is the mass density. The universal gas constant R (8.314 J K^{-1} mol^{-1}) is related to a specific gas constant (different for each atmosphere) by $\overline{R} = R/\overline{M} = K/\overline{m}$, where $\overline{M}$ (kg mol^{-1}) is the mean molar mass. Dalton's law states that the total pressure is the sum of partial pressures p_i, for each species i, $p = \sum p_i$. Applied to eqn [2], the mean molecular mass is

$$\overline{m} = \frac{\sum_i n_i m_i}{n} = \frac{\overline{M}}{N_A} \quad [3]$$

where m_i is the mass of the ith molecule and N_A is Avogadro's number. On Mars, for example, using the data of Table 2, the mean molar mass $\overline{M} = 0.0435\,\text{kg mol}^{-1}$ and $\overline{R} = R/\overline{M} = 191\,\text{J kg}^{-1}\,\text{K}^{-1}$. For dry air on Earth, $\overline{M} = 0.029\,\text{kg mol}^{-1}$ and $\overline{R} = R/\overline{M} = 287\,\text{J K}^{-1}\,\text{kg}^{-1}$.

We usually describe atmospheric composition with the volume mixing ratio f_i, for species i, which is also the mole fraction. Less commonly, we use a mass mixing ratio μ_i, which is the mass of a particular gas divided by the total mass of a gas mixture sample. For an ideal gas,

$$f_i = \frac{p_i}{p} = \frac{V_i}{V} = \frac{n_i}{n}, \quad \mu_i = \frac{\rho_i}{\rho}\frac{m_i}{\overline{m}}\frac{p_i}{p} = \frac{m_i}{\overline{m}}f_i \quad [4]$$

In all atmospheres, high total pressure at low altitudes pushes upward against the weight of overlying air, so that pressure declines with altitude. Air settles into hydrostatic equilibrium. This equilibrium assumes no vertical motion – a good approximation over large horizontal scales (e.g., a few kilometers on the Earth) that average out updrafts and downdrafts. A column of air with unit area cross section and small height Δz has a mass $\rho\Delta z$, so that the pressure change going up in height Δz is $\Delta p = -(\rho\Delta z)g =$ force/(unit area), where g is the gravitational acceleration. Written in calculus form, we get the hydrostatic equation relating pressure, density, and altitude:

$$\frac{\partial p}{\partial z} = -\rho(z)g(z), \quad p(z) = \int_z^\infty \rho(z)g(z)\mathrm{d}z \quad [5]$$

For an ideal gas, the air density is $\rho = p/\overline{R}T$ (eqn [2]), so that we can rewrite eqn [5] as

$$\frac{\partial p}{\partial z} = \left(\frac{-g(z)}{\overline{R}(z)T(z)}\right)p \Rightarrow \frac{\partial p}{p} = -\left(\frac{\partial z}{H}\right) \Rightarrow p = p_s\exp\left(-\int_0^z\left(\frac{\mathrm{d}z}{H}\right)\right) \Rightarrow p \approx p_s\exp\left(-\frac{z}{H}\right)_{\text{isothermal}} \quad [6]$$

Here, p_s is the surface pressure and we introduce a vertical length scale called the scale height,

$$H = \overline{R}T/g = kT/\overline{m}g \quad [7]$$

If we make the approximation that the temperature, gravity, and specific gas constant are constant with height, the integral of eqn [6] can be evaluated to give the exponential shown in the last expression. Thus, we see the meaning of the scale height: the vertical distance over which pressure drops by a factor of $1/e = 1/(2.7183) \approx 1/3$. In an isothermal atmosphere, density also drops by $1/e$ over the same height, but the density scale height will be slightly different if there is any temperature change with altitude. In the solar system, $H \approx 5$–20 km because the ratio $\overline{T}/\overline{m}g$ in $H = kT/\overline{m}g$ is similar given that smaller T (and larger g for Jupiter) compensates for small $\overline{m}$ in H_2-rich giant planets.

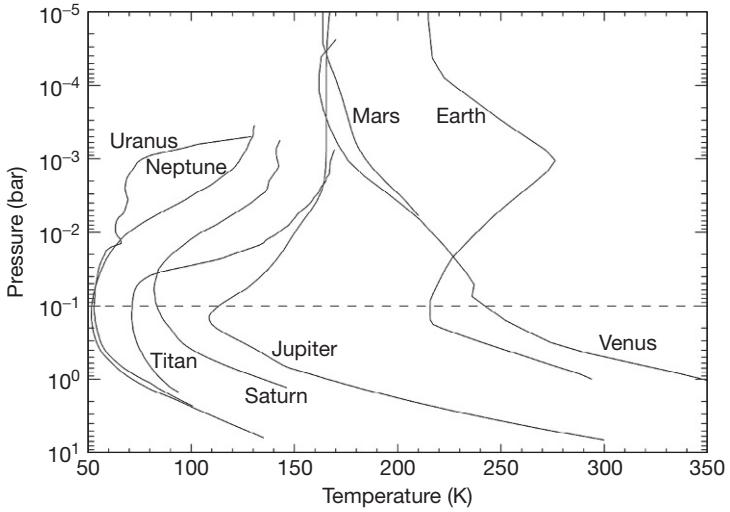


Figure 2 Temperature–pressure profiles for planetary atmospheres in the solar system. The dashed line at 0.1 bar indicates that planets with stratospheric temperature inversions have tropopause minima near this pressure level. (Drawn by the author.)

The surface pressure measures an atmosphere's mass because pressure is the weight of a column of air per unit area. Ignoring lateral forces in the real 3-D spherical geometry (which are a minor correction), we have

$$p_s = M_{col} g \Rightarrow M_{col} = \frac{p_s}{g} \qquad [8]$$

where M_{col} is the column mass. For example, $M_{col} = (610\ \text{Pa})/(3.72\ \text{m s}^{-2}) = 164\ \text{kg m}^{-2}$ on Mars, $10^4\ \text{kg m}^{-2}$ on Earth, and $2 \times 10^5\ \text{kg m}^{-2}$ on Titan. Equation [8] can also be applied to the column mass above some reference pressure level for giant planets.

The hydrostatic equation combined with number density gives us another way to describe atmosphere composition, namely, column abundance N_c, which is the number of molecules above a unit area. For example, the typical abundance of terrestrial ozone is $N_{O3} \approx 8 \times 10^{18}$ molecules cm^{-2}. At altitude z, the column abundance $N_c(z)$ is defined by

$$N_c(z) = \int_z^\infty n(z)\mathrm{d}z \qquad [9]$$

The column mass is $N_c(z)\overline{m}$, so the pressure at altitude z, by eqn [8], is $p(z) = N_c(z)\overline{m}g$. Using this, we obtain an expression for the column abundance in terms of scale height H:

$$N_c(z) = \frac{p(z)}{\overline{m}g} = \frac{n(z)kT(z)}{\overline{m}g} \Rightarrow N_c(z) = n(z)H(z) \qquad [10]$$

In some planetary literature, column abundance is expressed as 'cm-atm' or 'cm-amagat,' which are equivalent. These units are the depth in centimeters that would be produced if a gas in a column of air were all extracted and compressed to 1 atm (101325 Pa) pressure and 0 °C; they are given by the normalization $N_c(z)/n_0$, where $N_c(z)$ is in units of molecules cm^{-2} and $n_0 = 2.687 \times 10^{19}$ molecules cm^{-3} is the Loschmidt number, the number density of an ideal gas at 1 atm and 273.15 K. Ozone column abundance is also expressed with the Dobson unit (DU), which is the depth in thousandths of centimeters of a column of ozone taken to 1 atm and 0 °C, that is, DUs are $1000 N_{O3}/n_0$. The Earth's ozone column is typically 300 DU (0.3 cm-atm) compared with a maximum of 6 DU on Mars.

10.13.2.4 Convection and Stability

Atmospheric temperatures vary with altitude (Figures 1 and 2) primarily because of radiation and convection, although large-scale atmospheric motion also plays a role (Section 10.13.5).

In lower tropospheres, warm air rises and cold air sinks so that heat is transferred upward by convection. Air expands and becomes buoyant, that is, less dense than surroundings, through (i) radiative heating, (ii) contact with a planetary surface warmed by sunlight, (iii) exothermic condensation of substances such as water vapor in the Earth's atmosphere (see Chapter 10.11), and (iv) heating from the interior of a planet, which is significant on giant planets but not usually on rocky ones.

Convection commences when the lapse rate exceeds a certain threshold. The lapse rate is denoted $\Gamma = -(\mathrm{d}T/\mathrm{d}z)$ and usually given in K km^{-1}. The minus sign is a convention to make Γ positive for typical tropospheric conditions. Convection is very efficient in transporting heat upward to the upper troposphere. There, the overlying atmosphere is relatively transparent to thermal IR and energy is radiated to space. Also, in the upper troposphere, the lapse rate is no longer steep and convection ceases. In this way, a state of radiative–convective equilibrium is achieved, in which outgoing IR radiation from the warm planet balances the energy of incoming sunlight and sets a characteristic radiative–convective temperature profile (see Section 10.13.3.4).

The adiabatic lapse rate is the threshold lapse rate for convection. If the actual lapse rate is bigger or smaller, the atmosphere convects or it does not. Adiabatic means no exchange of energy with surroundings. So if a vertically displaced air parcel expands or contracts adiabatically to adjust to its surrounding pressure, the temperature change as a function of height is readily calculable. Applying conservation of energy to the parcel, we have

$$\underbrace{\mathrm{d}q}_{\text{heat added/kg (zero for adiabatic)}} = \underbrace{c_v \mathrm{d}T}_{\text{change in internal heat/kg}} + \underbrace{p\mathrm{d}\alpha}_{\text{work done/kg}} \qquad [11]$$

Here, dq is heat input per unit mass (zero if adiabatic) and c_v is the specific heat at constant volume. Work is 'force × incremental distance' or 'pressure (p) × incremental volume change.' Because we are dealing with energy per unit mass, we define volume per unit mass as $\alpha = 1/\text{density} = 1/\rho$, so that work is $p d\alpha$. Since $\alpha = \overline{R}T/p$ from eqn [2], the product rule of differentiation gives us $d\alpha = (\overline{R}/p)\,dT - (\overline{R}T/p^2)\,dp$. Substituting in eqn [11], we have

$$dq = c_v\,dT + \not{p}\left[\frac{\overline{R}}{\not{p}}\,dT - \frac{\overline{R}T}{p^{\not{2}}}\,dp\right] = (c_v + \overline{R})\,dT - \left(\frac{\overline{R}T}{p}\right)dp$$

$$\Rightarrow dq = c_p\,dT + g\,dz \qquad [12]$$

In the middle expression, we have inserted $c_p = c_v + \overline{R}$ from kinetic theory, which relates specific heat at constant pressure (c_p) to specific heat at constant volume. We also substituted $-g\,dz$ for $(\overline{R}T/p)\,dp$ using the first expression in eqn [5], the hydrostatic equation. For adiabatic changes, $dq = 0$. Thus, from the last expression in eqn [12], we obtain the adiabatic lapse rate for dry air:

$$\Gamma_a = -\left(\frac{dT}{dz}\right) = \frac{g}{c_p} \qquad [13]$$

This simple but fundamental equation gives the temperature change with altitude of a dry parcel of air moving up and down through an atmosphere in hydrostatic equilibrium. Table 3 gives the specific heat capacities, gravitational accelerations, and lapse rates Γ_a of various planetary atmospheres.

Comparison of the ambient lapse rate Γ with Γ_a indicates susceptibility to convection. When $\Gamma < \Gamma_a$, a parcel that rises adiabatically cools more than the gentler ambient decrease of temperature with height. This cold, dense parcel sinks back downward, so the atmospheric profile is convectively stable. Conversely, when $\Gamma > \Gamma_a$, a parcel that rises adiabatically is warmer than surroundings, and the parcel continues to rise. The resulting rapid mixing by convection of such unstable atmospheres tends to adjust the temperature profile toward the adiabatic one. Thus, we find planetary atmospheres typically with $\Gamma \leq \Gamma_a$ (ignoring local circumstances such as thunderstorms) (Table 3).

Our analysis is incomplete because condensable species and aerosols strongly affect temperature profiles. For example, in the Earth's troposphere, the condensation of water vapor releases heat making Γ smaller than the adiabatic value. Condensation of sulfuric acid on Venus has a similar thermodynamic effect. We discuss a saturated adiabatic lapse rate (SALR) in Section 10.13.2.5. For deep atmospheres, such as on Venus, one should also take account of the temperature dependence of specific heat capacity c_p so that $\Gamma_a = \Gamma_a(T)$.

A shorthand description of the stability of temperature gradients uses potential temperature θ, which is the temperature that a parcel of air would have if it were taken to a reference pressure p_{ref} (e.g., 10^5 Pa for Earth) dry adiabatically. Potential temperature is an alternative thermodynamic variable to temperature and, from its definition, is constant with altitude z if the lapse rate is adiabatic. Otherwise, an increase of θ with z indicates convective stability, while negative $\partial\theta/\partial z$ indicates instability. Using the adiabatic version of eqn [12] with $dq = 0$, we define θ as follows:

$$\frac{dT}{T} = \left(\frac{\overline{R}}{c_p}\right)\frac{dp}{p} = \kappa\frac{dp}{p};\quad \text{integrate } (\theta, T),\ (p_{ref}, p)$$

$$\Rightarrow \ln\left(\frac{T}{\theta}\right) = \kappa \ln\left(\frac{p}{p_{ref}}\right) \Rightarrow \theta = T\left(\frac{p_{ref}}{p}\right)^{\kappa} \qquad [14]$$

where $\kappa = \overline{R}/c_p$. The final expression in eqn [14] is Poisson's (adiabatic state) equation. We use eqn [14] later in Section 10.13.3.4, and potential temperature comes up again in Section 10.13.5.

In convectively stable atmospheres, a buoyancy force causes vertically displaced parcels to return to their original altitude, but overshooting creates oscillations at an angular frequency called the Brunt–Väisälä frequency, N_B. Large-scale oscillations of air create buoyancy waves (also called gravity waves) that are important in the turbulence of upper atmospheres (Section 10.13.5.7). The Brunt–Väisälä frequency can be derived by considering an adiabatic air parcel of volume V displaced upward δz from an altitude where the temperature is T to a place where the environmental temperature is T_{env}. The parcel cools adiabatically to a new temperature $T_p = T - \Gamma_a \delta z$, while the environment cools less at the ambient lapse rate as $T_{env} = T - \Gamma \delta z$. The difference in density between the parcel

Table 3 Thermodynamic and physical properties of various planetary atmospheres

Body	*Main gases*	*Molar mass*, M *(g mol^{-1})*	*Specific gas constant* $\overline{R}$ *(J kg^{-1} K^{-1})*	*Gravity* g *(m s^{-2})*	*Specific heat capacity* c_p *(J kg^{-1} K^{-1})*	$\Gamma_a = g/c_p$ *troposphere (K km^{-1})*	*Mean Γ observed (K km^{-1})*
Venus	CO_2, N_2	43.5	189	8.901	930	9.5	~8.0
Earth	N_2, O_2	28.97	287	9.81	1004	9.8	~6
Mars	CO_2, N_2	43.5	191	3.72	850	4.4	~2.5
Jupiter	H_2, He	2.22	3745	24.25	10,988	2.1–2.45[a]	1.9
Saturn	H_2, He	2.14	3892	10.0	10,658	0.7–1.1[a]	0.85
Titan	N_2, CH_4	28.67	290	1.36	1044	1.3	1.0–1.4
Uranus	H_2, He	2.3	3615	8.80	8643	0.7–1.1[a]	0.75
Neptune	H_2, He	2.3	3615	11.1	8643	0.85–1.34[a]	0.95
Hot Jupiter HD209458b	H_2, He	2.0	4160	18.5	14300	1.3	~0.2 (model)[b]

Γ_a is the tropospheric adiabatic lapse rate, while Γ is the observed tropospheric lapse rate.

[a]On giant planets, g varies from equator to pole.

[b]Menou and Rauscher (2009).

(ρ_p) and air (ρ_e) drives a buoyancy force, which we can write as Newton's second law:

$$\underbrace{(\mathcal{V}\rho_p)}_{\text{mass}}\underbrace{\frac{d^2(\delta z)}{dt^2}}_{\text{acceleration}}=\underbrace{-g(\rho_p-\rho_e)\mathcal{V}}_{\text{buoyancy force}}\quad\Rightarrow\frac{d^2(\delta z)}{dt^2}=g\left(\frac{\rho_e}{\rho_p}-1\right)=g\left(\frac{T_{env}}{T_p}-1\right)\qquad[15]$$

Substituting for temperatures T_p and T_{env}, we get

$$\frac{d^2(\delta z)}{dt^2}=g\left(\frac{T-\Gamma_a\delta z}{T-\Gamma\delta z}-1\right)=g\left(\frac{\Gamma\delta z-\Gamma_a\delta z}{T-\Gamma\delta z}\right)\approx g\left(\frac{\Gamma-\Gamma_a}{T-\cancel{\Gamma\delta z}}\right)\delta z$$

$$\Rightarrow\ddot{\delta z}+\frac{g}{T}(\Gamma_a-\Gamma)\delta z=0\qquad[16]$$

In the last expression of eqn [16], we recognize simple harmonic motion, that is, the '$\ddot{x}+\omega^2x=0$' of elementary physics. Thus, the square of the angular (Brunt–Väisälä) frequency of oscillation is

$$N_B{}^2\equiv\frac{g}{T}(\Gamma_a-\Gamma)\equiv\frac{g}{T}\left(\frac{dT}{dz}+\frac{g}{c_p}\right)\equiv\frac{g}{T}S\qquad[17]$$

Here, $S(\mathrm{K\,km^{-1}})$ is static stability, which measures the difference between the adiabatic lapse rate (eqn [13]) and actual lapse rate ($-dT/dz$) and is usually positive. Equation [17] makes intuitive sense: the more stable the atmosphere, the higher the oscillation frequency, like stiffening an oscillating spring.

The effect of buoyancy oscillations can be observed. Vertical deflection of horizontal air flow over an obstacle such as a mountain on Earth or a large crater rim on Mars sometimes causes oscillations where the air swings up and down beyond the obstacle at a frequency $\sim N_B/2\pi$, forming a standing buoyancy wave. At the cold crests of the waves, moisture can condense into a series of parallel clouds called lee wave clouds, given that they form in the lee of the obstacle. Roughly speaking, the distance between parallel clouds has wavelength, $\lambda=$(horizontal wind velocity/frequency)$=u/(N_B/2\pi)$, which obviously can change with altitude-dependent u and N_B. Approximate terrestrial values of $N_B=0.01\ \mathrm{s^{-1}}$ and $u=10\ \mathrm{m\,s^{-1}}$ give $\lambda\sim6$ km, consistent with satellite imagery of lee wave cloud spacing of 3–20 km.

10.13.2.5 Condensables and the Clausius–Clapeyron Equation

Gases that condense – condensables – affect atmospheres through release or uptake of heat when they change into a liquid or solid from the vapor phase, and vice versa. Condensables include H_2SO_4 (SO_3+H_2O) on Venus, H_2O on Earth, H_2O and CO_2 on Mars, H_2O and NH_3 on Jupiter and Saturn, and CH_4 on Titan, Uranus and Neptune, and HCN on Titan. Some condensables also form rain (e.g., CH_4 on Titan), snow (e.g., CO_2 in Mars' polar atmosphere), and clouds (e.g., CH_4 on Titan, Uranus, and Neptune).

The key concept governing condensable behavior is saturation vapor pressure (SVP), which is the vapor pressure exerted in equilibrium with a liquid or solid condensed phase. The curve of SVP versus temperature for a particular substance defines the transition between vapor and condensed phases

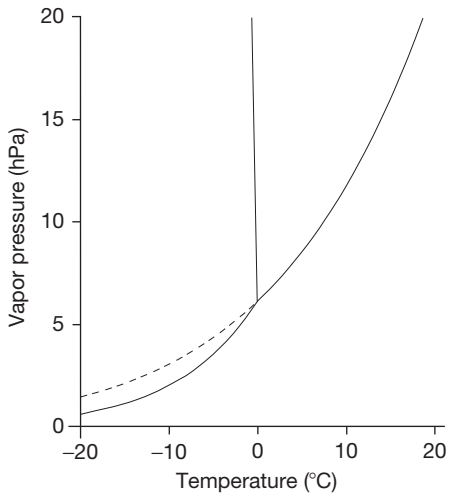


Figure 3 The saturation vapor pressure (SVP) curve between the vapor phase and condensed phase (liquid or ice) for water. The dashed line is the SVP with supercooled water. (Drawn by the author.)

(Figure 3). The Clausius–Clapeyron equation defines this SVP curve as a unique function of temperature. On Earth, we use a Clausius–Clapeyron equation for water, but analogous equations apply for other condensables. When noncondensable gases are part of a gas mixture, the vapor pressure of the condensable gas is a partial pressure, and if that partial pressure reaches SVP, then the solid or liquid phase forms.

The Clausius–Clapeyron equation comes from thermodynamics. Two phases are in equilibrium along the SVP versus T line, so the Gibbs free energy change of the phase transition is zero everywhere along the line. If e_s is the SVP, it can be shown by considering Gibbs free energy that the slope of e_s versus temperature T is given by (e.g., Atkins and De Paula, 2010)

$$\frac{de_s}{dT}=\frac{\delta s}{\delta\alpha}=\frac{\text{entropy gained per kg from condensed to vapor}}{\text{volume increase per kg from condensed to vapor}}=\frac{l_c}{T\delta\alpha}=\frac{l_c}{T(\alpha_{vapor}-\alpha_{condensed})}\approx\frac{l_c}{T\alpha_{vapor}}\qquad[18]$$

where α is specific volume ($\alpha=1/\rho$), which, in the last step, we reasonably assumed, is far larger than for the vapor than for condensed phase, that is, $\alpha_{vapor}\gg\alpha_{condensed}$. We also substituted $l_c=T\delta s$, where l_c is the latent heat or enthalpy, which is heat absorbed (or released) when a unit mass changes from one phase to another at constant temperature and pressure. For different phase transitions, this quantity can be 'latent heat of vaporization' for liquid–vapor, or latent heat of fusion for ice–liquid, or latent heat of sublimation for ice–vapor. We can define a specific gas constant for the condensable species vapor as $R_c=R/M_c=k/m_c$, where M_c and m_c are its molar mass and molecular mass, respectively. By substituting for $\alpha_{vapor}=R_cT/e_s$ from the ideal gas law, we obtain the Clausius–Clapeyron equation:

$$\frac{de_s}{dT}=\frac{l_ce_s}{R_cT^2}\Rightarrow\frac{d(\ln e_s)}{dT}=\frac{l_c}{R_cT^2}\Rightarrow e_s(T)=e_s(T_0)\exp\left(\int_{T_0}^{T}\frac{l_c}{R_c}\frac{dT}{T^2}\right)\approx e_s(T_0)\exp\left[\frac{l_c}{R_c}\left(\frac{1}{T_0}-\frac{1}{T}\right)\right]\qquad[19]$$

Here, the last step makes the approximation that l_c does not change with temperature. Because exponents are additive,

eqn [19] shows $e_s \propto \exp(-m_c l_c/kT)$. Thus, SVP is a Boltzmann equation, which makes sense: the numerator $m_c l_c$ is the energy to free a molecule from its neighbors, while the denominator kT is the average molecular energy. For water vapor above liquid water, we use $T_0 = 273.15\,\mathrm{K}$, $e_s(T_0) = 611\,\mathrm{Pa}$, $R_c = 461\,\mathrm{J\,K^{-1}\,kg^{-1}}$, and latent heat of vaporization $l_c = 2.5 \times 10^6\,\mathrm{J\,kg^{-1}}$, finding that SVP roughly doubles for every 10 K temperature rise.

On some bodies, SVP is essential for understanding the bulk atmospheric gas. On Mars, for example, CO_2 condenses at the cold poles. A temperature of 148 K corresponds to an SVP of 6 mbar (~600 Pa) for CO_2, which is a typical air pressure on the surface of Mars (Leighton and Murray, 1966). Consequently, Mars's CO_2 ice caps are buffered at a temperature near 148 K. On Triton and Pluto, the atmospheric pressure is essentially the SVP of N_2 over N_2 ice at temperatures of 38 and ~40 K, respectively (Brown and Ziegler, 1979; Lellouch et al., 2011; Zalucha et al., 2011).

The presence of even a minor condensable can change the lapse rate considerably, as water vapor does on Earth. Once condensation starts in a cooled, vertically displaced parcel of air, further condensation (into liquid or ice) reduces the rate of cooling by release of latent heat. An SALR, or moist adiabat, denoted Γ_{sa}, occurs when chilled moist air is held at saturation by the warming associated with continual condensation of vapor. This lapse rate is called the pseudoadiabatic lapse rate if condensates are assumed to fall out of the parcel because then the process is not reversible and adiabatic. However, saturated and pseudoadiabats tend to be similar because condensation products carry relatively little heat out of a parcel of air. The SALR can be derived as a function of temperature and saturation mass mixing ratio $\mu_{cms} = (m_c/\overline{m}) f_c = \epsilon(e_s/p)$, where f_c is the volume mixing ratio of the condensable, p is the total pressure, and ϵ is the ratio of molecular masses. In this way, the SALR can be expressed as a modification of the adiabatic lapse rate, that is, $\Gamma_{sa} = \Gamma_a \times f(T, \mu_{cms})$ (e.g., see Andrews (2010), pp. 37–39). Formulations of the SALR for major constituent condensation are also possible (Ingersoll, 1969; Kasting, 1988, 1991).

Under saturated conditions, static stability is controlled by comparing the actual lapse rate to the saturated lapse rate, Γ_{sa}. Air is stable if $\Gamma < \Gamma_{sa}$. For example, on Earth, $\Gamma_{sa} \approx 4\,\mathrm{K\,km^{-1}}$ in warm humid air near the ground, $\Gamma_{sa} \approx 6$–$7\,\mathrm{K\,km^{-1}}$ in the middle troposphere, and Γ_{sa} is close to dry adiabatic in the cold, dry upper troposphere. In the Earth's tropics, observations show that moist convection controls static stability up to ~0.4 bar (Zelinka and Hartmann, 2011).

A further nuance applies to giant planets, where the weight of saturated air parcels is greater than the dry atmosphere unlike the case for water-saturated parcels on Earth or methane-saturated parcels on Titan that tend to be buoyant. Ammonia and water on Jupiter and Saturn have greater molecular mass (17 and $18\,\mathrm{g\,mol^{-1}}$, respectively) than the hydrogen–helium dry atmospheres (2.22 and $2.14\,\mathrm{g\,mol^{-1}}$, respectively), and the same situation applies to methane on Uranus and Neptune. If the abundance of a condensable species drops faster in the environment with altitude than in a rising air parcel, the parcel will tend to be heavy relative to the environment and sink back downward. Consequently, convection on giant planets is favored when the relative humidity in the environment is high and disfavored by a steep compositional gradient in the environment due to condensation. Temperature profiles on Uranus and Neptune are superadiabatic where methane condenses, which suggests that convection may be inhibited there (Guillot, 1995).

SVP also gives us a first-order understanding of clouds. Clouds will form on average where the mean pressure–temperature conditions reach the SVP for various condensables (Sanchez-Lavega et al., 2004) (**Figure 4**).

10.13.3 Radiation

10.13.3.1 Planetary Energy Balance, Blackbody Radiation, and the Greenhouse Effect

Physical and chemical processes in planetary atmospheres are driven by sources of free energy, meaning energy that can do useful work. For most planets, light from the parent star dominates free energy sources, although some giant planets have an internal energy flux that leaks to the exterior and is significant compared to the energy flux received from the parent star.

Consequently, for planetary atmospheres, the solar flux at the top of the atmosphere is a key parameter. Satellites measure the Earth's solar constant as $S_0 = 1360.8 \pm 0.5\,\mathrm{W\,m^{-2}}$ (Kopp and Lean, 2011), consistent with ground-based inferences (Chapman et al., 2012). As a result, the solar flux at the top of the atmosphere averaged over the area of the whole Earth and over time is $S_0/4 \approx 340\,\mathrm{W\,m^{-2}}$. We divide S_0 by 2 to account for glancing angles on a spherical planet, and we divide it by 2 again because only one side of the Earth is in daylight. Alternatively, another way to understand the factor of ¼ is that the solar flux intercepted by a planet of radius R is equivalent to that on a projected disk of area πR^2. So, the ¼ factor represents the ratio of received flux on a disk to the whole planetary area ($\pi R^2/4\pi R^2$) to get a global average in $\mathrm{W\,m^{-2}}$.

A fraction of sunlight's incident power is reflected and scattered back to space, which, over all wavelengths, is called the Bond albedo A (or planetary albedo). For the solar system planets with atmospheres, A ranges from 0.25 for Mars to 0.76 for Venus (Moroz et al., 1985), the giant planets have $A \approx 0.3$, while for Earth, A is 0.3 (Trenberth et al., 2009). Thus, the solar flux absorbed by the Earth's surface and atmosphere is $(1-A) \times (S_0/4) = (1\text{–}0.3) \times (1361\,\mathrm{W\,m^{-2}}/4) \approx 238\,\mathrm{W\,m^{-2}}$.

An essential concept for understanding what happens to the energy on planets is blackbody radiation. All objects emit a spectrum of electromagnetic radiation based on their temperature (e.g., a molten lava emits visible light plus radiation extending into the IR), and for practical purposes, blackbody radiation can be simply considered as an idealized spectrum that depends only on the temperature of the emitting object. Temperature determines the flux of emitted radiation as a function of wavelength, and when the flux is summed over all wavelengths, the total flux depends only on temperature also. Historically, the blackbody concept originated in considering radiation within a closed cavity that had walls that perfectly absorbed and emitted radiation at all wavelengths. At equilibrium, the spectrum was found to be uniquely related to the cavity wall temperature T by the Planck function, $B(T)$:

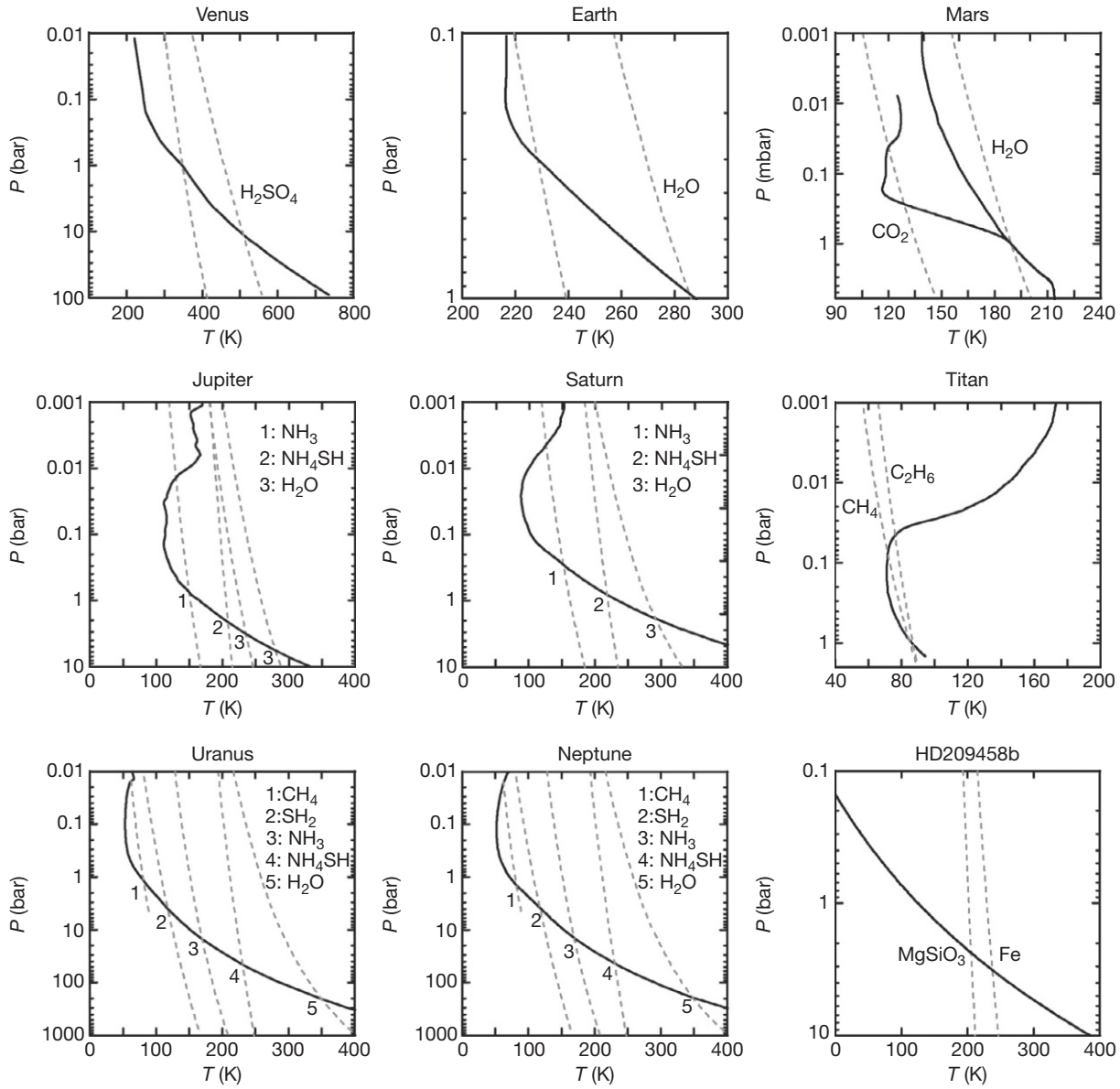


Figure 4 Use of the vapor pressure and saturation vapor pressure (SVP) for various condensables to predict cloud formation in solar system atmospheres and hot Jupiter HD209458b. The solid line is the vertical profile of pressure (p) versus temperature (T). For Mars, an annual average p-T profile is shown as well as a colder profile. The dashed lines are the saturation vapor pressure (SVP) curves for various condensable species, as indicated, assuming a fixed volume mixing ratio of condensable. It is assumed that particles condense when the partial pressure (e) reaches the saturation vapor pressure (e_s), that is, where $e = f_c p(T) e_s(T)$, where f_c is the volume mixing ratio of the condensable. Thus, the base of clouds is marked where a dashed line crosses the solid pressure–temperature profile. Below this level, the phase is vapor because the ambient air is too hot. Various mixing ratios are assumed to calculate the dashed curves as follows: Venus 2 ppm and 2000 ppm H_2SO_4 (where liquid clouds form); the Earth 250 ppm and 1.5% H_2O, giving rise to ice and liquid clouds, respectively; Mars 0.95 CO_2 and 300 ppm H_2O giving ice clouds; Jupiter 200 ppm NH_3, 36 pppm NH_4SH, 50 ppm H_2O, 0.17% H_2O giving solid phase clouds except for the higher water level; Saturn 200 ppm NH_3, 36 pppm NH_4SH, 0.17% H_2O, giving solid particles; Titan 5% CH_4, 10 ppm C_2H_6, giving solid particles; Uranus and Neptune 2% CH_4, 37 ppm SH_2, 200 ppm NH_3, 36 pppm NH_4SH, 0.17% H_2O, giving solid particles; and HD209458b 75 ppm $MgSiO_3$ and 68 ppm Fe, giving solid particles. Reprinted from Sanchez-Lavega A, Perez-Hoyos S, and Hueso R (2004) Clouds in planetary atmospheres: A useful application of the Clausius–Clapeyron equation. *American Journal of Physics* 72: 767–774.

$$\underbrace{B_\upsilon \,\mathrm{d}\upsilon = \frac{2h\upsilon^3}{c^2}\frac{\mathrm{d}\upsilon}{(\mathrm{e}^{h\upsilon/kT}-1)}}_{B_\upsilon \text{ in } \mathrm{Wm^{-2}\,Hz^{-1}\,sr^{-1}}},\quad \underbrace{B_\lambda \,\mathrm{d}\lambda = \frac{2hc^2}{\lambda^5}\frac{\mathrm{d}\lambda}{(\mathrm{e}^{hc/\lambda kT}-1)}}_{B_\lambda \text{ in } \mathrm{Wm^{-2}\,m^{-1}\,sr^{-1}}},$$

$$\underbrace{B_{\tilde{\upsilon}}\,\mathrm{d}\tilde{\upsilon} = 2hc^2\frac{\tilde{\upsilon}^3\,\mathrm{d}\tilde{\upsilon}}{(\mathrm{e}^{hc\tilde{\upsilon}/kT}-1)}}_{B_{\tilde{\upsilon}} \text{ in } \mathrm{Wm^{-2}\,(cm^{-1})^{-1}\,sr^{-1}}} \qquad [20]$$

Here, $h = 6.626 \times 10^{-34}$ J s is Planck's constant and c is the speed of light, which is related to the frequency (υ) and wavelength (λ) of radiation by $c = \upsilon\lambda$. The third expression in eqn [20] is in terms of wavenumber $\tilde{\upsilon} = 1/\lambda$, as used by some spectroscopists. Thus, each form of the Planck function in the preceding text applies when the spectrum is specified in units of frequency (Hz), wavelength (m), or wavenumber (usually $\mathrm{cm^{-1}}$), respectively. In eqn [20], blackbody radiation $B_\upsilon(T)$ is a

radiance (or intensity), which is the power (W) per unit frequency (or wavelength or wave number) interval crossing unit area (m^{-2}) in unit solid angle (steradian^{-1} or sr^{-1}) in a given direction.

A flux density or irradiance (often shortened to just 'flux') has units of $Wm^{-2}\,Hz^{-1}$ (or $Wm^{-2}\,m^{-1}$ or $Wm^{-2}\,cm^{-1}$) and is the intensity integrated over all directions. In spherical polar coordinates, an elemental solid angle is $\sin\theta d\theta d\phi$ where θ is the angle from the polar axis and ϕ is the azimuthal angle. If the intensity B_υ is isotropic (i.e., independent of θ), the flux over an entire hemisphere is π times B_υ, as follows:

$$F_\upsilon = \int_0^{\phi=2\pi}\int_0^{\theta=2\pi} B_\upsilon \cos\theta \underbrace{\sin\theta}_{x}\, d\theta d\phi = 2\pi\int_0^{x=1} B_\upsilon x\, dx$$
$$= 2\pi B_\upsilon \left[\frac{x^2}{2}\right]_0^1 = \pi B_\upsilon \qquad [21]$$

The flux over all frequencies ($W\,m^{-2}$) is found by substituting for B_υ (eqn [20]) in eqn [21] and integrating

$$F = \int_0^\infty \pi B_\upsilon\, d\upsilon = \frac{2\pi h}{c^2}\int_0^\infty \frac{\upsilon^2 d\upsilon}{\left(e^{h\upsilon/kT}-1\right)}$$
$$= \frac{2\pi k^4 T^4}{c^2 h^3} \underbrace{\int_0^\infty \frac{x^3 dx}{(e^x-1)}}_{\text{standard integral}=\pi^4/15} = \left(\frac{2\pi^5 k^4}{15c^2h^3}\right)T^4 \Rightarrow F = \sigma T^4 \qquad [22]$$

Here, we substituted $x = h\upsilon/kT$ to obtain a standard integral. The final expression in eqn [22] is the Stefan–Boltzmann law with the Stefan–Boltzmann constant $\sigma = 5.67\times10^{-8}\,W\,m^{-2}\,K^{-4}$.

Blackbodies are ideal absorbers and emitters, and because planets are made of materials that absorb IR radiation strongly and reach thermal equilibrium, the spectrum that planets emit can be represented by a blackbody. Essentially, a planet absorbs visible sunlight, warms up, and radiates like a blackbody in the IR. When outgoing radiation balances incoming radiation along with an additional internal heat flux F_i (as occurs on the giant planets), the planet achieves an effective temperature T_e, that is, the equivalent blackbody temperature that characterizes the total flux output:

$$\underbrace{(1-A)\frac{S_p}{4}}_{\text{absorbed solar radiation}} + \underbrace{F_i}_{\text{internal heat flux}} = \underbrace{\sigma T_e^4}_{\text{outgoing IR radiation}},$$

$$\text{or } (1-A)\frac{S_p}{4} = \underbrace{\sigma T_{eq}^4}_{\text{outgoing IR radiation}} \quad \text{if } F_i \ll (1-A)\frac{S_p}{4} \qquad [23]$$

Here, S_p is the solar flux at the orbital distance of the planet from the Sun. When the internal heat flux is negligible, the effective temperature is called the equilibrium temperature, T_{eq}, because there is equilibrium between incoming sunlight and outgoing IR. On Jupiter, F_i is $\sim5.4\,W\,m^{-2}$ compared with $8.1\,W\,m^{-2}$ of absorbed sunlight (Hanel et al., 1981). For giant planets, sources of internal energy include gravitational energy released from formation and contraction as well as subsequent internal differentiation. However, unlike the other gas giants, Uranus has little internal heat flow ($42\pm47\,mW\,m^{-2}$ (Pearl et al., 1990)), which is puzzling. For the Earth, $F_i \sim 0.09\,W\,m^2$ can be neglected, $S_p = 1361\,W\,m^{-2}$, and Bond albedo $A = 0.3$. Solving eqn [23] gives equilibrium temperature $T_{eq} = 255$ K, which is also the Earth's effective temperature. Clearly, this temperature is not achieved at the surface, which has a mean temperature of 288 K. For planets with atmospheres, the difference between the surface temperature and effective temperature arises because of the greenhouse effect described in the succeeding text.

An idealized atmospheric level where Earth achieves its effective temperature can be estimated. The Earth's average tropospheric lapse rate is $\sim6\,K\,km^{-1}$, so given the surface temperature of 288 K, the air temperature will be at 255 K at ~5 km altitude. This altitude is called the emission level, although it is somewhat fictitious because in reality, the atmosphere radiates to space from different levels as a function of wavelength, depending upon the IR transparency of the overlying air at that wavelength. Nonetheless, the Earth cannot buck conservation of energy, and in equilibrium, it must radiate a total flux back to space equivalent to a 255 K blackbody of the Stefan–Boltzmann law.

The greenhouse effect is when an atmosphere warms up and provides a flux of thermal IR radiation to the surface (or lower layers of giant planets) that adds to the shortwave solar flux. The surface of a rocky planet with an atmosphere is warmer than it would be without an atmosphere because the surface receives energy from both the Sun and the heated atmosphere. For specificity, the fraction of the total flux reaching the surface due to IR from the atmosphere is about 99.9%, 2/3, 1/3, and 9/10 for Venus, Earth, Mars, and Titan, respectively (Courtin et al., 1992). For rocky planets, the difference between T_{eq} and surface temperature T_s is another way to express the size of the greenhouse effect:

$$\Delta T_g \equiv T_s - T_{eq} \qquad [24]$$

For the Earth, $\Delta T_g = 288 - 255 = 33$ K. Of course, this definition assumes that the albedo would be the same where the air removed – which it certainly would not – so ΔT_g should be viewed judiciously as a useful metric, a way to express the flux difference of about $150\,W\,m^{-2}$ between blackbodies at 255 and 288 K. Of course, slowly rotating airless bodies can have large day–night temperature differences, and so, the mean temperature will be different from the equilibrium temperature, as is well known (Pierrehumbert, 2010).

There are various greenhouse gases. In the Earth's modern clear sky atmosphere, water vapor contributes $\sim$60% of ΔT_g, CO $\sim$25%, and ozone $\sim$8%, with the rest from CH_4, N_2O, anthropogenic chlorofluorocarbons (CFCs), and other trace gases (Karl and Trenberth, 2003; Stephens and Tjemkes, 1993). Importantly, water vapor condenses and exits the atmosphere as rain or snow, whereas the Earth's other main greenhouse gas, CO_2, does not condense. H_2O is effectively 'a slave' to the forcing by the concentration of CO_2. By forcing, we mean a persistent disturbance to the climate system that changes the energy balance. If CO_2 levels goes up, more water vapor enters a warmer atmosphere and increases the greenhouse effect. Thus, the climatic role of water vapor is a feedback rather than a forcing (Lacis et al., 2010, 2013). The bulk background gases that absorb very little IR, such as N_2 and O_2, are themselves

critical for the greenhouse effect because collisions of greenhouse gases (CO_2, H_2O, etc.) with N_2 or O_2 take away or add energy to the greenhouse gas molecules, which allows them to absorb over a much wider range of photon wavelengths. This process of pressure broadening (see **Section 10.13.3.5**) is fundamental to the greenhouse effect in thick atmospheres. On other bodies, rather than just trace gases, the bulk atmospheric gases are greenhouse gases. On all the giant planets, H_2 is the main source of IR opacity (see **Section 10.13.3.5**), on Titan, N_2, H_2, and CH_4 are important (McKay et al., 1991; Samuelson et al., 1981), while CO_2 is the key greenhouse gas on Venus (Bullock and Grinspoon, 2001; Titov et al., 2007) and Mars.

Clouds can cool a planet because of their contribution to the albedo, but they can also warm through adding to the greenhouse effect. For example, on Earth, icy cirrus clouds contribute to greenhouse warming since they are largely transparent to solar radiation but opaque to thermal IR, while low-lying marine stratus predominantly contributes to shortwave albedo. If we take clouds into account in our apportioning of the greenhouse effect on Earth, water vapor is responsible for ~50%, clouds ~25%, CO_2 ~20%, and other gases the remainder (Schmidt et al., 2010). Regarding their cooling effect, clouds contribute about half of the Earth's albedo of 0.3. According to satellite measurements, clouds reduce the Earth's absorbed radiation by $-48\,W\,m^{-2}$, but enhance the greenhouse effect by $+30\,W\,m^{-2}$. So, overall, clouds cool the Earth by $-18\,W\,m^2$ (Ramanathan and Inamdar, 2006). A hypothetical Earth without clouds would be far warmer (Goldblatt et al., 2013).

Like water vapor, clouds are a feedback and change in response to levels of CO_2 and other greenhouse gases. Feedbacks can alter the mean temperature of a planet in positive feedback (like the aforementioned water vapor feedback) or stabilize it in negative feedback. Another important positive feedback is the ice-albedo feedback. An increase in surface temperature causes a decrease in albedo from diminished snow and ice cover, which causes further warming, or the other way, cooling increases albedo with more snow and ice, causing further cooling. The ice-albedo feedback is crucial for understanding the Earth's ice ages during the last few million years (Lorius et al., 1990). Negative feedbacks include an increase in outgoing thermal IR as a planet warms up, operating on short timescales of weeks. Over geologic timescales, geochemical cycles act as negative feedbacks on the Earth's climate. The important one for Earth is the carbonate–silicate negative feedback, which acts over a timescale of ~0.5 My (Berner, 2004; Walker et al., 1981). If the climate warms, greater rainfall and temperatures consume atmospheric CO_2 in weathering reactions on the continents, which cools the Earth. If the climate cools, the consumption of CO_2 from dry air is slowed, and CO_2 accumulates due to volcanic emissions, which increases the greenhouse effect. On long times, negative feedbacks have evidently 'won' overall because the Earth has remained habitable since at least 3.5 Ga (Buick, 2007).

One further importance of blackbody theory is that it tells us the wavelength where most of the radiative energy is emitted from planets and stars. The Planck function is maximal where $\partial B_\lambda/\partial\lambda = 0$, which gives Wien's law:

$$\lambda_{max} T = 2.897 \times 10^{-3}\ \text{meters} \bullet K \qquad [25]$$

The Sun behaves like a blackbody because the ions and electrons inside the Sun are in thermal equilibrium. So, solar radiation has an effective temperature of 5780 K. Wien's law indicates peak emission at $\lambda_{max} = (0.0029\,m\,K)/(5780\,K) = 0.5\,\mu m$, in the visible. The Earth is roughly 20 times cooler than the Sun's surface, and so, its Wien peak is 20 times larger at ~10 μm. Given typical temperatures, planets generally have their Wien peak at thermal IR wavelengths, which is where many atmospheric gases and aerosol particles absorb strongly.

Several atmospheres have windows, which are spectral intervals that are relatively transparent to thermal IR compared to high opacity at other wavelengths. A window can be a key means for a planet to radiate IR energy to space and cool. On Titan, a window at 16.5–25 μm is close to Titan's Wien peak of 31 μm. H_2, which is a minor constituent (~0.1%), absorbs weakly in this window (McKay et al., 1989). On Earth, the key atmospheric window of 8–12 μm encompasses the Wien peak at ~10 μm. Again, absorption by a minor constituent, water vapor, modulates the outgoing IR flux through the window. On Venus, dense CO_2 and clouds leaves only a few narrow windows in the near-IR. The significant ones are 1–1.2 μm in which only H_2O absorbs, 1.7 μm where CO and HCl absorb, and 2.2–2.5 μm where CO and SO_2 are seen (Crisp and Titov, 1997; Titov et al., 2007). These are wavelengths to peer relatively deep into Venus' atmosphere, which is otherwise opaque.

10.13.3.2 Radiative Transfer: Shortwave and Infrared

Atmospheres like the Earth's are transparent to incoming solar radiation but not entirely. The ozone in the Earth's stratosphere absorbs solar UV (0.2–0.4 μm) in the Hartley and Huggins bands, and ozone in both the stratosphere and troposphere absorbs visible light in the 0.4–0.85 μm Chappuis band. The Earth's tropospheric water vapor absorbs solar near-IR (1–4 μm region). A more extreme case is Titan's stratospheric methane and haze of aerosols, which absorb in the UV, visible, and near-IR (for methane) to such an extent that the reduction of solar flux reaching Titan's surface is called an antigreenhouse effect (McKay et al., 1989, 1991). Temporary antigreenhouse effects are also possible from debris raised in asteroid impacts (Pollack et al., 1983) or, on Earth, smoke created in nuclear war (Robock et al., 2007; Turco et al., 1983).

Radiative transfer restricted to the visible and UV is simpler than in the IR because the Planck function (eqn [20]) indicates that we can neglect visible or UV emission at temperatures typical of planetary surfaces and atmospheres. This is not true for IR calculations. The starting point in radiative transfer is the extinction law (or Beer–Lambert–Bouguer law). Experiments show that the loss of intensity of a beam of light of frequency ν passing through a substance depends linearly on the incident intensity I_ν and the elemental mass of absorbing or scattering material expressed as $\rho_a\,ds$, where ρ_a is the density of the absorber or scatterer ($kg\,m^{-3}$) and ds is the elemental path length:

$$dI_\nu = -k_\nu I_\nu \rho_a ds \Rightarrow I_\nu = I_{\nu 0}\exp\left(-\int_0^s k_\upsilon \rho_a\, ds'\right) = I_{\nu 0} e^{-\tau_\nu} \qquad [26]$$

Here, the proportionality constant k_υ is the mass extinction coefficient ($m^2\,kg^{-1}$) or opacity. In eqn [26], $I_{\nu 0}$ is the intensity at distance $s = 0$ and τ_ν is a dimensionless optical path or

optical thickness along the beam direction. We usually define our optical path as increasing from the top of the atmosphere downward. In this geometry, we call τ_v the optical depth. In the case of a pencil beam inclined at zenith angle θ (from the vertical), the elemental path along the beam ds is related to that along the vertical path dz by geometry as $\cos\theta = dz/ds$. Replacing ds in eqn [26] by $dz/\cos\theta$ gives

$$I_v = I_{v0} e^{-\tau_v/\cos\theta} \Rightarrow \text{transmissivity } T_v = e^{-\tau_v/\cos\theta} \quad [27]$$

Here, the transmissivity (or transmittance) T_v is the ratio of light getting through (I_v) to the incident intensity (I_{v0}).

The visible optical depth can be large in planetary atmospheres. On Mars, the broadband visible optical depth in global dust storms reaches ~5 (Colburn et al., 1989; Kahn et al., 1992), which means that direct-beam sunlight is attenuated by a factor of $e^{-5} \approx 1/148$. On Titan's surface at 0.5 μm visible wavelength, the extinction optical depth is ~10 due to atmospheric haze (Tomasko and West, 2009). However, these surfaces are not dark because the total flux at the ground is the sum of a direct flux from the Sun's direction and a diffuse flux from the whole sky. The diffuse flux includes all scattered solar radiation from the sky, including scattered light from clouds or particles that is not in the direct beam. On Mars, for example, much light scattered by dust out of the direct solar beam reaches the ground diffusely (Haberle et al., 1993a). On Earth, when the sky is covered by dense cloud and the Sun cannot be seen, the ground is similarly well illuminated by a diffuse flux.

The extinction coefficient k_v considered in the preceding text is the sum of absorption and scattering due to gases or aerosols. We define the ratio of scattering to extinction as single scattering albedo ω_0,

$$k_v = \underbrace{a_v}_{\text{absorption coefficient}} + \underbrace{s_v}_{\text{scattering coefficient}}, \quad \omega_0 = \frac{s_v}{a_v + s_v} \quad [28]$$

The single scattering albedo is 1 for purely scattering media and <0.5 for strongly absorbing media. It is one of the three key parameters used to characterize optical properties of aerosols or cloud particles. The second is a volume absorption (or scattering) coefficient defined as $Q_{as}A_{xs}n$, where Q_{as} is a dimensionless absorption or scattering efficiency (when absorption and scattering are combined, Q_{as} can be redefined as an extinction efficiency), A_{xs} is the geometric cross section, and n is number density. Thus, we would write $dI_v = -I_v Q_{as} n A_{xs} ds$ in eqn [26]. For terrestrial cloud droplets, $Q_{as} \sim 2$ in the visible. The third parameter is the asymmetry factor g_s, a function that indicates whether scattering is isotropic ($g_s = 0$) or scatters in the forward ($0 < g_s \leq 1$) or backward ($-1 \leq g_s < 0$) direction. These parameters are used in models of scattering; for example, see Petty (2006).

Absorption of photons warms aerosols, causing them to heat atmospheres during daytime (e.g., dust on Mars (Pollack et al., 1979) or black carbon on Earth (Ramanathan and Carmichael, 2008)), while for gases, absorption can excite a molecule or atom into vibration or rotation or both. At high pressures, collisions generally occur before an excited atom or molecule undergoes radiative decay and emits a photon in a random direction, with the result that vibrational or rotational energy is converted into KE and heat, called thermalization or quenching. Absorption of EUV photons in the upper atmosphere can instead create ions in photoionization or break molecules in photodissociation.

In contrast to absorption, scattering is when radiation is redirected out of the original direction of propagation as a result of interaction with aerosols or when radiative decay of a molecule emits a photon in a random direction, attenuating the direct beam. The type of scattering can be diagnosed from the size of a scatterer using a dimensionless size parameter, where r_p is the scatterer radius and λ is the wavelength of radiation,

$$x = \frac{2\pi r_p}{\lambda} \quad [29]$$

For gas molecules or particles with $r_p \ll 1$, the scattering of solar radiation is Rayleigh scattering, which is isotropic with a scattering efficiency $Q_s \propto \lambda^{-4}$, a strong wavelength dependence that causes the Earth's blue sky when convolved with the solar spectrum. Aerosols (dust, smoke, and cloud droplets) with $0.1 < x < 50$ scatter according to Mie scattering (for spherical particles), typically with forward scattering. Mie theory gives the electromagnetic field around a dielectric particle. Photons can be scattered into the beam and scattered out of the beam in multiple scattering by aerosols. Finally, for particles with $x > 50$, there is just geometric scattering, where $Q_s \approx 2$ and the scattering cross section grows in proportion to the cross-sectional area πr_p^2.

To deduce information about the atmospheric structure and composition, we frequently view an IR spectrum from a planet remotely. On Earth, satellite measurements show that our planet does not have the smooth idealized curve of the Planck function but has big 'dips' in the spectrum compared to a blackbody curve at ~320 K surface temperature (in the case of **Figure 5(a)**) because of absorption by key greenhouse gases. The bottom of the dips indicates the effective temperature of the emission where photons are emitted not from the surface but a cooler atmospheric level, for example, the bottom of the CO_2 band at 15 μm emits at about the tropopause temperature, while water vapor at 20 μm emits from around the mid-troposphere (**Figure 5(b)**). The area under the Planck curves gives the total power emitted, which is reduced by the area of the dips. Thus, in order to maintain equilibrium between solar flux received and planetary flux emitted, the surface temperature (which is seen in the atmospheric window at 8–12 μm) increases above a 255 K blackbody curve. Effectively, **Figure 5(a)** visualizes the Earth's greenhouse effect.

To make a quantitative sense of remote measurements such as those in **Figure 5(a)**, we use the physics of IR radiative transfer. If we consider a monochromatic beam of radiance I_v [W m^{-2} Hz^{-1} sr^{-1}] passing through an elemental path ds, then the intensity change will be that emitted into the beam minus that extinguished by absorption or scattering:

$$dI_v(s) = \underbrace{(k_v \rho_a ds) J_v}_{\text{light scattered in or emitted}} - \underbrace{(k_v \rho_a ds) I_v}_{\text{extinction}} \Rightarrow \frac{dI_v}{k_v \rho_a ds} = J_v - I_v \quad [30]$$

Here, ρ_a is the density of the absorbing or scattering gas and J_v [W m^{-2} Hz^{-1} sr^{-1}] is a source function, which accounts for thermal emission or scattering into the beam. The term labeled 'extinction' uses the extinction law, eqn [26]. The final expression in eqn [30] is the general radiative transfer equation with

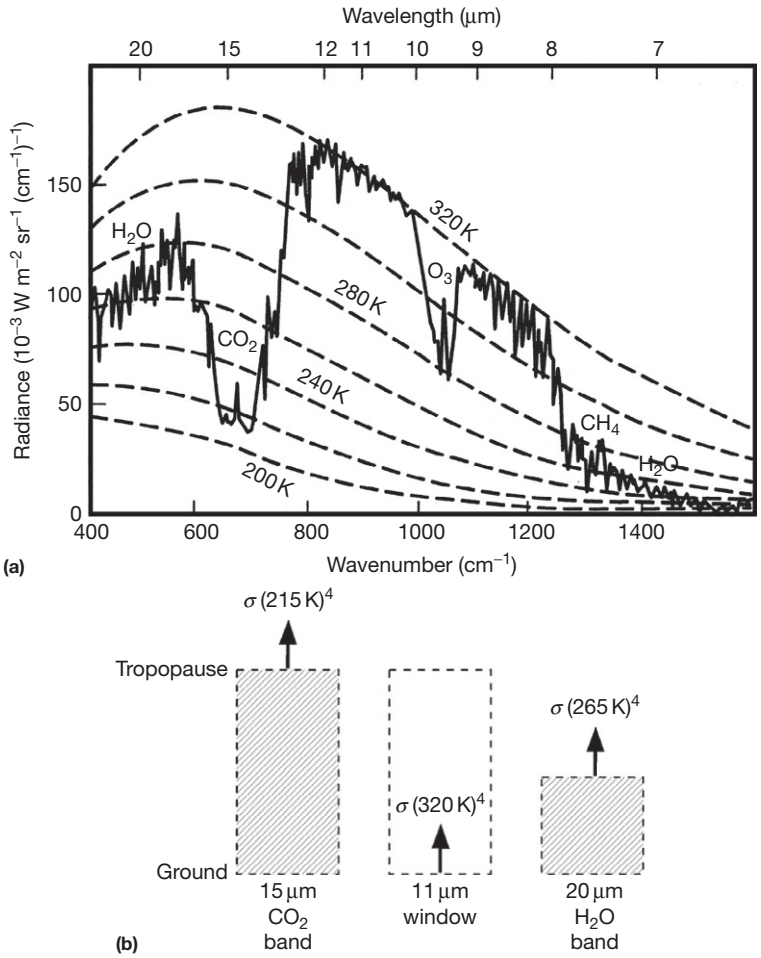


Figure 5 (a) The Planck function for noon over a vegetated region of the Niger Valley in N. Africa where the surface temperature is 320 K. Dashed lines show blackbody curves for particular temperatures (Adapted from Hanel RA, Conrath BJ, Kunde VG, et al. (1972) The Nimbus 4 infrared spectroscopy. Experiment, I: calibrated thermal emission spectra. *Journal of Geophysical Research* 77: 2629–2641). (b) A schematic diagram showing the meaning of parts of the curve in (a). The arrows indicate the layers in the atmosphere where blackbody fluxes originate, according to the Stefan–Boltzmann law. (Drawn by the author.)

no particular coordinate system or assumption about whether the source function behaves like a blackbody or not.

Next, we impose a coordinate system and assume that the medium is a blackbody, that is, a purely absorbing, non-scattering medium in thermal equilibrium with source function $J_\nu = B_\nu =$ the Planck function and k_ν the mass absorption coefficient. Minus signs are confusing in radiative transfer due to different geometric conventions, so let us be clear that we now define our IR optical depth as measured downward along a path s, that is, τ is zero at an observer at $s = \infty$ and increases backward along the beam to some value τ inside the atmosphere. Thus,

$$\tau = \int_s^\infty k_\nu \rho_a \, ds = \int_{\tau=\tau}^{\tau=0} -d\tau, \quad \text{or} \quad \text{if } s \text{ is altitude:}$$

$$\tau = \int_z^\infty k_\nu \rho_a \, dz = \int_{\tau=\tau}^{\tau=0} -d\tau \qquad [31]$$

With these assumptions, the equation of radiative transfer becomes

$$\frac{dI_\nu(s)}{d\tau_\nu(s_1, s)} = I_\nu(s) - B_\nu(s) \qquad [32]$$

where the elemental optical path $d\tau$ is between distance s and some distance s_1. This is sometimes called Schwarzschild's equation after Karl Schwarzschild who first considered such a radiative transfer equation for the Sun's atmosphere in 1914. With a little work, a solution to eqn [32] can be found by multiplying by an integrating factor, $e^{-\tau_\nu(s_1, s)}$, integrating between 0 and s_1, and rewriting as

$$I_\nu(s_1) = I_{\nu 0} e^{-\tau_\nu(s_1, 0)} + \int_0^{s_1} B_\nu(s) e^{-\tau_\nu(s_1, 0)} k_\upsilon \rho_a \, ds \qquad [33]$$

The first term is just the exponential attenuation of the emission from $s = 0$ to $s = s_1$. The second is the sum of the contributions from each point s along the path 0 to s_1, each exponentially attenuated by the path length to the observer.

Another useful form of the radiative transfer equation makes the plane-parallel approximation, that is, where pressure, temperature, and composition are assumed to be functions of height only. Then, if a beam of radiance I_ν passes upward

through a layer of depth dz at an angle θ to the vertical, geometry gives $dz = ds\cos\theta = (ds)\mu$, defining the cosine of zenith angle as $\mu = \cos\theta$. Given our definition of optical depth (eqn [31]), the equation for radiative transfer in a plane-parallel atmosphere is a slight modification of eqn [32], as follows, where we return to generality by using J_ν:

$$\mu\frac{dI_\nu}{d\tau_\nu} = I_\nu - J_\nu \qquad [34]$$

If we imagine looking down on a planet from some level in the atmosphere, we integrate eqn [34] from that level where the optical depth is τ to the surface where $\tau = \tau_*$. Given τ as a limit of integration, we use a dummy variable $\tau' - \tau$ in the integral for path lengths. To obtain the upward intensity at any level where the optical depth is τ, we multiply by an integrating factor $e^{-(\tau'-\tau)/\mu}$ and solve. Assuming that a υ (frequency) subscript on τ is implicit, the solution for upward intensity is

$$I_\nu(\tau)^\uparrow = I_\nu(\tau_*)e^{-(\tau_*-\tau)/\mu} + \int_\tau^{\tau_*} J_\nu(\tau')e^{-(\tau'-\tau)/\mu}\frac{d\tau'}{\mu} \quad \left(1 \geq \mu > 0\right) \qquad [35]$$

where the arrow superscript indicates upward. If we are looking down from space, we simply put $\tau = 0$ in the limit of integration, and the intensity observed at the satellite instrument is

$$I_\nu(\tau=0)^\uparrow = \underbrace{I_\nu(\tau_*)e^{-\tau_*/\mu}}_{\text{surface contribution}} + \underbrace{\int_0^{\tau_*} J_\nu(\tau')e^{-\tau'/\mu}\frac{d\tau'}{\mu}}_{\text{contributions along the path}} \qquad [36]$$

Here, the first term is the extinction along the path from the ground to the spacecraft where $I_\nu(\tau_*)$ is the upward intensity at the ground. The second term is the upward emission contribution from points along the path to the spacecraft. Here, $J_\nu(\tau')$ corresponds to emission in the upward direction. The same expression can also be written in altitude coordinates but is more complicated:

$$I_\nu(z)^\uparrow = I_\nu(z=0)e^{-\frac{1}{\mu}\int_0^z k_\nu(z')\rho_a(z')dz'} + \int_0^z k_\nu(z')\rho_a(z')J_\nu(z')e^{-\frac{1}{\mu}\int_0^z k_\nu\rho_a dz''}\frac{dz'}{\mu} \qquad [37]$$

We derive a hemispheric monochromatic flux F_ν from the measured intensity I_ν at a particular frequency ν by integrating over a hemisphere. Then, we derive a total flux (W m^{-2}) by integrating over all frequencies. We will use eqn [37] in the succeeding text.

10.13.3.3 Optical Thickness and Thinness

Vertical sections of atmospheres are described as optically thick ($\tau_\nu \gg 1$) or optically thin ($\tau_\nu \ll 1$) to radiation of a particular frequency. 'Optically thick' just means opaque to a particular frequency or frequency band of radiation, and 'optically thin' means nearly transparent. This is a useful concept in looking at spectra such as **Figure 5(a)**. The IR photons can be thought of as 'escaping' from the level where the IR optical depth drops below ~1 and then passing up through overlying 'optically thin' or IR-transparent air to space. We can also apply the idea to incoming solar flux. Solar photons are mainly absorbed in the atmosphere near the level of unity optical depth for a particular UV or visible wavelength. As a result, UV heating or ionization can occur at this level.

It is possible to justify the concept of absorption near $\tau_\nu \sim 1$ quantitatively for both solar and IR radiation. For solar radiation, we work with fluxes directly because of the typical lack of emission in atmospheres at solar frequencies. Within a planetary atmosphere at height z, the vertical component of direct flux crossing a surface parallel to the ground for a solar beam at zenith angle θ and frequency ν can be deduced from the extinction law (eqn [27]) as

$$F_\upsilon = \mu F_{\upsilon\odot}e^{-\tau_\upsilon/\mu} \qquad [38]$$

The volume heating rate Q [W m^{-3} Hz^{-1}] is related to how this flux changes with altitude due to absorption by the expression, $Q = -(dF_{\text{net},\nu}/dz)$, where $F_{\text{net},\nu}$ is the net flux, that is, the upward minus the downward flux. With no upward solar flux, the heating rate is

$$Q = \frac{dF_\upsilon}{dz} = \underbrace{\frac{dF_\upsilon}{d\tau}}_{-k_\upsilon\rho_\upsilon}\frac{d\tau_\upsilon}{dz} = \left(-\frac{1}{\mu}\mu F_{\upsilon\odot}e^{-\tau_\upsilon/\mu}\right)(-k_\upsilon\rho_\upsilon) \Rightarrow Q = (k_\upsilon\rho_\upsilon F_{\upsilon\odot})e^{-\tau_\upsilon/\mu} \qquad [39]$$

Initially, the value of Q in eqn [39] increases downward as the term $k_\nu\rho_\nu$ in the curved brackets increases because τ_ν is small at the top of the atmosphere and the exponential factor is near 1. Eventually, the optical depth becomes significant, so the term exp $(-\tau_\upsilon/\mu)$ begins to dominate over the downward increase of $k_\nu\rho_\nu$, and Q begins to decrease exponentially downward toward zero. We deduce that Q has a single maximum at some altitude in between.

We find the maximum in Q by differentiating w.r.t. z, using the product rule and setting $dQ/dz = 0$

$$\frac{dQ}{dz} = 0 = \left(k_\upsilon\rho_\upsilon\right)\left(\cancel{F_{\upsilon\odot}e^{-\tau_\upsilon/\mu}}\frac{\left(k_\upsilon\rho_\upsilon\right)}{\mu}\right) + \cancel{F_{\upsilon\odot}e^{-\tau_\upsilon/\mu}}\frac{d\left(k_\upsilon\rho_\upsilon\right)}{dz}$$
$$\Rightarrow -\frac{d\left(k_\upsilon\rho_\upsilon\right)}{dz} = \frac{\left(k_\upsilon\rho_\upsilon\right)^2}{\mu} \Rightarrow \frac{d^2\tau_\nu}{dz^2} = \frac{1}{\mu}\left(\frac{d\tau_\nu}{dz}\right)^2 \qquad [40]$$

In the last step, we use $d\tau_\upsilon/dz = -k_\upsilon\rho_\upsilon$ from the definition of optical depth (eqn [31]). Because the density of absorbing gases usually drops exponentially with increasing height and the absorption coefficient is either constant or an approximate power law relationship with pressure, optical depth should vary approximately exponentially with altitude, that is, $\tau_\upsilon = \tau_{\upsilon 0}$ exp (z/H_τ) where H_τ is a scale height for the optical depth. If this expression for τ_υ is inserted into eqn [40], after a little manipulation, an important result is obtained that $dQ/dz = 0$ occurs where

$$\tau_\upsilon \sec\theta = 1, \quad \text{or} \quad \tau_\upsilon = 1 \quad \text{when zenith angle } \theta = 0 \qquad [41]$$

Thus, at any frequency, the maximum volume heating rate Q, or related effect such as dissociation or ionization, occurs where the optical depth is unity for normal incidence radiation. The altitude is higher by factor sec θ for solar radiation incident at zenith angle θ. Also, the characteristic width of activity is $\sim H_\tau$. Such a vertical structure is called a 'Chapman layer'.

A similar insight about the importance of unity optical depth applies in the thermal IR for atmospheres that are

optically thick at depth, that is, when we cannot see the surface emission looking downward from space. In our expression for upward intensity from an atmosphere (eqn [37]), we neglect the first term due to surface radiation, extend the integral to the whole atmosphere, and assume blackbody emission. Then,

$$I_\nu(\infty)\uparrow \approx \sec\theta \int_0^\infty k_\nu \rho_a B_\nu \exp\left[-\sec\theta \int_z^\infty k_\nu \rho_a dz'\right] dz$$
$$\Rightarrow I_\nu(\infty)\uparrow = \sec\theta \int_0^\infty B_\nu W_\nu \, dz \qquad [42]$$

where W_ν is a weighting function, defined by

$$W_\nu(z) \equiv k_\nu(z)\rho_a(z)\exp\left[-\sec\theta \int_z^\infty k_\nu(z')\rho_a(z')dz'\right] \qquad [43]$$

Equation [42] says that the radiance to space is a weighted vertical integral of the Planck function. The weighting function W_υ is mathematically similar to heating rate Q in eqn [39]. Just like Q, function W_υ is zero at the bottom of the atmosphere where the exponential reaches zero, W_υ is zero at the top of the atmosphere where density is zero, and W_υ has a maximum in between. As a result, the monochromatic IR radiance comes from a layer with characteristic thickness that is a scale height in optical depth or smaller and centered where optical depth has exactly the same $\tau_\upsilon \sim 1$ condition as eqn [41]. For frequencies with large $k_\upsilon \rho_a$ that are highly absorbing, the $\tau_\upsilon \sim 1$ peak of the weighting function W_υ will be high up (large z), given that τ_υ is the integral of $k_\upsilon \rho_a$ from z to ∞. This is seen in a planetary IR spectrum such as **Figure 5(a)** in dips where greenhouse gases absorb. The dependence of $I_\nu(\infty)\uparrow$ on gas composition and wavelength in eqn [42] is also how we deduce the temperature and chemical structure of planetary atmospheres using remote sensing, for example, from orbiters around other planets or ground-based telescopes.

10.13.3.4 Radiative–Convective Equilibrium

Combining the theory of convection (**Section 10.13.2.4**) and radiation (**Sections 10.13.3.1–10.13.3.3**) allows us to understand the average vertical structure of planetary atmospheres quantitatively given that it is primarily controlled by these processes.

A hypothetical atmospheric structure determined only by radiation balance (ignoring convection) will generate an unrealistic temperature profile in an IR-opaque troposphere with a steep temperature gradient that is actually unstable to convection. Conceptually, a temperature profile for purely radiative transfer for a stratosphere and upper troposphere has to be joined to a convective profile below with both continuities of temperature and flux. How do we calculate this? For analytic calculations, three approximations are made, as follows:

- The first is a gray approximation, where the thermal opacity of the atmosphere is assumed to be independent of wavelength and represented by a single, broadband value. 'Gray opacity' should be viewed as a fitting parameter to match an observed thermal structure. It rarely makes sense to laboriously compute gray opacities from spectral averaging, and it is a misconception to consider gray opacities as 'real' properties of atmospheres, given that the real opacity is wavelength-sensitive. Instead, a gray optical property gives a characterization of an atmosphere that is useful for semi-quantitative comparison of planetary atmospheres.
- The second is that the diffuse approximation is implemented. This replaces radiances I and B in Schwarzschild's radiative transfer equation with fluxes, with the optical depth scaled by a factor D, called the diffusivity factor to account for the integration of the radiances over a hemisphere. A value of $D = 1.66$ is close to numerical results (Armstrong, 1968; Rodgers and Walshaw, 1966) and equivalent to assuming an average zenith angle of 53°, that is, $1/\mu = \sec\theta = 1.66$, in eqn [34]. Others set $D = 3/2$ (Goody and Yung, 1989; Weaver and Ramanathan, 1995) or $D = 2$.
- The third simplification is to handle thermal IR with a two-stream approximation, in which upward (F^+) and downward (F^-) fluxes are treated separately.

Radiative–convective equations are easiest to formulate with optical depth τ as the vertical coordinate but can be converted into physical space using a relationship between τ and pressure. In a plane-parallel atmosphere, Schwarzschild's equation (eqn [34]) for the upwelling and downwelling thermal radiative fluxes (F^+ and F^-, respectively) can be written as

$$\frac{dF^+}{d\tau} = D(F^+ - \pi B), \quad \frac{dF^-}{d\tau} = -D(F^- - \pi B);$$
$$\text{where } \pi B(\tau) = \sigma T^4(\tau) \qquad [44]$$

Equation [44] allows calculation of upwelling and downwelling thermal fluxes if $T(\tau)$ is specified. The net thermal flux is $F_{net} = F^+ - F^-$. Subtracting $dF^-/d\tau$ from $dF^*/d\tau$ in eqn [44], we get

$$\frac{dF_{net}}{d\tau} = D(F^+ + F^- - 2\pi B) \qquad [45]$$

and differentiating w. r. t. τ gives

$$\frac{d^2F_{net}}{d\tau^2} = D\left(\underbrace{\frac{d(F^+ + F^-)}{d\tau}}_{DF_{net}} - 2\pi\frac{dB}{d\tau}\right) \Rightarrow \frac{d^2F_{net}}{d\tau^2} - D^2F_{net} = -2\pi D\frac{dB}{d\tau} \qquad [46]$$

In the last step, addition of $dF^-/d\tau$ and $dF^+/d\tau$ from eqn [44] gives $d(F^+ + F^-)/d\tau = DF_{net}$. Equation [46] can be solved for blackbody radiance $B(\tau)$ and the atmospheric temperature profile using the Stefan–Boltzmann law if $F_{net}(\tau)$ and a boundary condition are known.

Boundary conditions involve the solar flux. In the solar system, atmospheres absorb solar radiation in the troposphere or at the surface and partly in stratospheres. The latter often leads to stratospheric temperature inversions (**Figure 2**), while absorption in tropospheres can stabilize upper tropospheres against convection, as happens on Titan. Exponential attenuation is a reasonable approximation for solar radiation, so the net absorbed stellar flux $F^\odot_{net}$ can be written as (Robinson and Catling, 2012)

$$F^\odot_{net}(\tau) = \underbrace{F^\odot_{net} e^{-k_{tropo}\tau}}_{\text{tropospheric (and surface) solar absorption}} + \underbrace{F^\odot_{strato} e^{-k_{stratio}\tau}}_{\text{stratospheric solar absorption}} \qquad [47]$$

Here, $F^\odot_{strato}$ and $F^\odot_{tropo}$ are the top-of-atmosphere net absorbed stellar fluxes in the stratosphere and troposphere (including

the surface for planets with surfaces), respectively. Parameters $k_{\text{strato}} = \tau_{\text{sws}}/\tau$ and $k_{\text{tropo}} = \tau_{\text{swt}}/\tau$ control the attenuation of solar flux in two channels, where τ_{sws} and τ_{swt} are shortwave optical depths and τ is the gray thermal IR optical depth. At the top of the atmosphere, where $\tau = 0$, the net absorbed stellar flux must be (eqn [23])

$$F^{\odot}_{\text{net}}(\tau = 0) = (1 - A)\frac{F^{\odot}}{4} \qquad [48]$$

where A is the planet's Bond albedo and $F^{\odot}$ is the solar flux incident at the top of the atmosphere. The overall energy balance at any atmospheric level is between the net thermal flux with the absorbed stellar flux and any internal energy flux, F_{i} from a planet's interior (relevant for giant planets):

$$F_{\text{net}}(\tau) = F^{\odot}_{\text{net}}(\tau) + F_{\text{i}} \qquad [49]$$

By combining eqn [47] with eqn [49] and inserting into eqn [46], we obtain

$$\frac{\mathrm{d}\sigma T^4}{\mathrm{d}\tau} = \frac{1}{2D}\left[\left(D^2 - k^2_{\text{tropo}}\right)F^{\odot}_{\text{tropo}}\mathrm{e}^{-k_{\text{tropo}}\tau} + \left(D^2 - k^2_{\text{strato}}\right)F^{\odot}_{\text{strato}}\mathrm{e}^{-k_{\text{strato}}\tau} + D^2 F_{\text{i}}\right] \qquad [50]$$

which integrates to

$$\sigma T^4(\tau) = \sigma T^4(0) + \frac{1}{2D}\left[\frac{D^2 - k^2_{\text{tropo}}}{k_{\text{tropo}}}\left(1 - \mathrm{e}^{-k_{\text{tropo}}\tau}\right)F^{\odot}_{\text{tropo}} + \frac{D^2 - k^2_{\text{strato}}}{k_{\text{strato}}}\left(1 - \mathrm{e}^{-k_{\text{strato}}\tau}\right)F^{\odot}_{\text{strato}} + D^2 F_{\text{i}}\tau\right] \qquad [51]$$

The flux at the top of the atmosphere in this equation ($\sigma T^4(0)$) can be eliminated by applying the boundary condition that there is no downwelling thermal radiation at the top of the atmosphere where $\tau = 0$. We first use the definition $F_{\text{net}} = F^{+} - F^{-}$ to rewrite eqn [45] as

$$\frac{\mathrm{d}F_{\text{net}}}{\mathrm{d}\tau} = D\left(F_{\text{net}} + 2F^{-} - 2\sigma T^4\right) \qquad [52]$$

and, second, we solve for σT^4 using eqns [49] and [47]:

$$\sigma T^4(\tau) = F^{-}(\tau) + \frac{1}{2}\left[\frac{D + k_{\text{tropo}}}{D}F^{\odot}_{\text{tropo}}\mathrm{e}^{-k_{\text{tropo}}\tau} + \frac{D + k_{\text{strato}}}{D}F^{\odot}_{\text{strato}}\mathrm{e}^{-k_{\text{strato}}\tau} + F_{\text{i}}\right] \qquad [53]$$

Third, inserting $\tau = 0$, and noting that $F^{-}(0)$ is zero, gives $\sigma T^4(0)$. Inserting $\sigma T^4(0)$ into eqn [51] gives a final expression for the temperature profile in the radiative part of a gray atmosphere that sits above a radiative–convective boundary and extends into the stratosphere:

$$\sigma T^4(\tau) = \frac{F^{\odot}_{\text{strato}}}{2}\left[1 + \frac{D}{k_{\text{strato}}} + \left(\frac{k_{\text{strato}}}{D} - \frac{D}{k_{\text{strato}}}\right)\mathrm{e}^{-k_{\text{strato}}\tau}\right] + \frac{F^{\odot}_{\text{tropo}}}{2}\left[1 + \frac{D}{k_{\text{tropo}}} + \left(\frac{k_{\text{tropo}}}{D} - \frac{D}{k_{\text{tropo}}}\right)\mathrm{e}^{-k_{\text{tropo}}\tau}\right] + \frac{F_{\text{i}}}{2}(1 + D\tau) \qquad [54]$$

If there are no internal heat ($F_{\text{i}} = 0$) and no solar attenuation (the limit $k_{\text{strato}} = k_{\text{tropo}} \to 0$, applying l'Hôpital's rule), eqn [54] reduces to $\sigma T^4(\tau) = (F^{\odot}_{\text{net}}(0)/2)(1 + D\tau)$, which is a common form for simple radiative models in textbooks (e.g., Andrews (2010), p. 85; Goody and Yung (1989), p. 392). Using $\sigma T^4(\tau)$ in eqn [54], the downwelling flux can also be deduced from eqn [53] and the upwelling flux from a similar expression.

We can now ask: what is required for an atmosphere to have a strong stratospheric temperature inversion and thus a well-developed tropopause minimum? The minimum occurs in the radiative regime, and we can simplify eqn [54] by omitting terms in tropospheric shortwave attenuation, that is, $k_{\text{tropo}} \ll 1$, which is the case for real atmospheres. Then, setting the derivative w.r.t. τ of the resulting expression to zero, the optical depth of the tropopause minimum is (Robinson and Catling, 2014)

$$\tau_{\text{tp}} = \underbrace{\frac{1}{k_{\text{strato}}}}_{\sim 1/100}\underbrace{\ln\left[\underbrace{\frac{F^{\odot}_{\text{strato}}}{F^{\odot}_{\text{tropo}} + F_{\text{i}}}}_{\sim 1/10}\underbrace{\left(\frac{{k_{\text{strato}}}^2}{D^2} - 1\right)}_{\sim 10^3}\right]}_{\sim 5} \qquad [55]$$

The numbers given under parts of eqn [55] are typical averages for the Earth, Titan, and the giant planets, which all have stratospheres and tropopause minima (**Table 4**). Using these numbers, the gray IR optical depth where the tropopause minimum occurs is $\tau_{\text{tp}} \sim 0.05$. This optical depth depends on $F^{\odot}_{\text{strato}}/(F^{\odot}_{\text{tropo}} + F_{\text{i}})$, which is the ratio of the stratospheric absorbed flux relative to that from below and k_{stratio}, which is the ratio of shortwave optical depth to IR optical depth in the stratosphere. Equation [55] only makes mathematical sense if

$$k^2_{\text{strato}} > D^2\left[1 + \left(F^{\odot}_{\text{tropo}} + F_{i}\right)/F^{\odot}_{\text{strato}}\right] \qquad [56]$$

which sets a threshold condition on k_{strato} for a tropopause minimum or, alternatively, the condition for a stratospheric temperature inversion. It tells us that to have a stratospheric inversion, an atmosphere has to have a large k_{strato}, that is, a strong shortwave absorber with a shortwave optical depth that dominates over the IR opacity; otherwise, cooling to space by the IR-active molecules cancels shortwave heating of a stratosphere. Typical flux ratios in the planets of the solar system in eqn [56] mean that k_{strato} must be $\sim 10^2$ for a well-developed tropopause minimum.

The condition for a stratospheric inversion (eqn [56]) is satisfied by some planets but not by others depending on stratospheric composition and the balance of shortwave absorption to IR cooling (**Table 5**). The Earth's stratospheric ozone absorbs strongly in the UV, so k_{strato} is large and heating is not overwhelmed by emission by CO_2 at 15 μm, water vapor, or ozone at 9.6 μm. Consequently, the Earth has a stratospheric inversion. In contrast, CO_2 in the atmospheres of Mars and Venus does not absorb strongly in the visible or solar near-IR; instead, in the upper atmosphere, the 15 μm band of CO_2 radiates heat to space efficiently. Thus, k_{strato} is too small for stratospheric inversions in the global average on Mars and Venus. On the giant planets, dark stratospheric aerosols absorb in the UV and visible, while methane has many

Table 4 Solar absorption and infrared properties of gray radiative–convective parameters for the Earth, Titan, and the giant planets

Parameter	*World*					
	Earth	Jupiter	Saturn	Titan	Uranus	Neptune
P_{ref} [bar]	1	1	1	1.5	1	1
T_0 [K]	288	166	135	94	76	72
α_c	0.6	0.85	0.94	0.77	0.83	0.87
$F_{strato}^{\odot}$ [W m^{-2}]	7	1.3	0.41	1.49	0.24	0.09
$F_{tropo}^{\odot}$ [W m^{-2}]	233	7.0	2.04	1.12	0.41	0.18
F_i [W m^{-2}]	Minor, 0.087	5.4	2.01	Minor	Minor, 0.042 ± 0.047	0.43
k_{strato}	90	90	180	120	220	580
k_{tropo}	0.16	0.06	0.03	0.20	0.08	0.20
τ_{rc}	0.15	0.34	0.44	4.4	0.62	0.41
τ_0	1.9	6.3	9.2	5.6[a]	8.7	3.0
τ_{tp}	0.050	0.064	0.040	0.077	0.042	0.017
p_{tp} [bar] (model)	0.16	0.10	0.066	0.18	0.070	0.075
p_{tp} [bar] (observed)	0.16	0.14	0.08	0.2	0.11	0.1

[a]For Titan, $\tau_0 = 2.5$ at $p = 1$ bar.

Parameters inferred from measurements include a tropospheric reference pressure (P_{ref}) and temperature (T_{ref}), an empirical adjustment factor applied to the dry adiabat to account for volatile condensation (α_c) (eqn [59]), an internal energy flux (F_i), and the solar flux absorbed in the stratosphere ($F_{strato}^{\odot}$) and troposphere ($F_{tropo}^{\odot}$). For rocky bodies, $F_{tropo}^{\odot}$ includes surface absorption. For example, the Earth has $F_{strato}^{\odot} + F_{tropo}^{\odot} = 240\,\mathrm{W/m^2}$, the global mean absorbed solar flux. Parameters that control solar energy attenuation in the stratosphere and troposphere are $K_{strato} = \tau_{sws}/\tau$ and $K_{tropo} = \tau_{swt}/\tau$, where τ_{sws} and τ_{swt} are shortwave optical depths and τ is the gray thermal IR optical depth at any atmospheric level (eqn [47]). The gray IR optical depths at the reference pressure, radiative–convective boundary, and tropopause are τ_0, τ_{rc} and τ_{tp}, respectively.

A comparison of τ_0 values shows that IR optical depth is somewhat larger on the giant planets than the Earth or Titan. Also, Titan has the largest fraction of solar radiation absorbed in its stratosphere.

Source: Robinson TD and Catling DC (2014) Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency *Nature Geoscience* 7: 12–15.

Table 5 Molecules that dominate stratospheric shortwave heating and infrared cooling in atmospheres of the solar system

Planet	*Stratospheric heating*	*Stratospheric cooling*
Oxidizing stratospheres		
Venus	Unknown absorber (UV), CO_2, sulfur-bearing aerosols	CO_2, sulfur-bearing aerosols
Earth	Ozone	CO_2, water vapor, ozone
Mars	CO_2	CO_2
Reducing stratospheres		
Jupiter	Aerosols (UV/vis), CH_4 (NIR)	H_2–H_2 and H_2–He CIA continuum, acetylene (C_2H_2), ethane (C_2H_6), methane
Saturn	Aerosols (UV/vis), CH_4 (NIR)	H_2–H_2 and H_2–He CIA continuum, acetylene (C_2H_2), ethane (C_2H_6), methane
Titan	Haze, CH_4 (NIR)	Acetylene (C_2H_2), ethane (C_2H_6), haze
Uranus	Aerosols (UV/vis), CH_4 (NIR)	H_2–H_2 and H_2–He CIA continuum, acetylene (C_2H_2), ethane (C_2H_6), methane
Neptune	Aerosols (UV/vis), CH_4 (NIR)	H_2–H_2 and H_2–He CIA continuum, acetylene (C_2H_2), ethane (C_2H_6), methane

CIA, collision-induced absorption.

absorption bands in solar near-IR. A collision-induced absorption (CIA) continuum of H_2–H_2 and H_2–He and thermal IR emission bands of acetylene (C_2H_2, 13.7 μm) and ethane (C_2H_6, 12.1 μm) and methane (7.6 μm) cool these stratospheres inefficiently (Yelle et al., 2001; Zhang et al., 2013). Overall, k_{strato} is large, and all the giant planets have stratospheric inversions. On Titan, a thick stratospheric haze absorbs visible and UV radiation, along with stratospheric absorption in the near-IR by methane. The results are a large k_{strato} and a pronounced stratospheric inversion.

The aforementioned is radiation only; the other part of the radiative–convective problem to consider is the convective regime. Equation [14] describes a dry adiabat, which we can rewrite as temperature as a function of pressure,

$$T = T_{ref}\left(\frac{p}{p_{ref}}\right)^{\overline{R}/c_p} \Rightarrow T = T_{ref}\left(\frac{p}{p_{ref}}\right)^{(c_p - c_v)/c_p} \Rightarrow T = T_{ref}\left(\frac{p}{p_{ref}}\right)^{(\gamma-1)/\gamma} \quad [57]$$

Here, T_{ref} is a reference temperature at a reference pressure p_{ref} (e.g., 1 bar), while γ is the ratio of specific heats c_p/c_v. Kinetic theory relates γ to the degrees of freedom, N_{dof}, for the primary atmospheric constituent(s), where

$$\gamma = 1 + \frac{2}{N_{dof}} \quad [58]$$

For atmospheres of linear diatomic gases, such as N_2–O_2 (Earth), H_2 (giant planets), or N_2 (Titan), the molecules have

3 translational and 2 rotational degrees of freedom, so that $N_{dof}=5$ and $\gamma=7/5=1.4$. For CO_2-dominated atmospheres (Venus and Mars), there are 3 translational, 2 rotational, and ~0.3 vibrational degrees of freedom, so $N_{dof}=6.3$ and $\gamma=1.3$ (Bent, 1965). With these γ values, a tropospheric temperature T is proportional to $p^{0.3}$ and $p^{0.2}$ for diatomic and triatomic planetary atmospheres, respectively. However, no observed troposphere follows a dry adiabat because condensables release latent heat and lower the lapse rate. Following Robinson and Catling (2012), a semiempirical approximation adjusts the convective tropospheric temperature structure as follows:

$$T = T_{ref}\left(\frac{p}{p_{ref}}\right)^{\alpha_c(\gamma-1)/\gamma} \quad [59]$$

where α_c is a dimensionless constant, ~0.6–0.9 for planets in the solar system, representing the average ratio of the observed lapse rate to dry adiabatic lapse rate in the planet's convective regime.

Equation [59] gives temperature as a function of pressure in the convective regime, whereas eqn [54] gives temperature as a function of optical depth in the radiative regime, so how is optical depth related to pressure? The gray differential optical depth, $d\tau$, across an elemental layer of vertical thickness dz is the sum of various contributions from N_a sources of opacity, such as absorbing gases:

$$d\tau = -\sum_{i=1}^{N_a} k_i \underbrace{\rho_i}_{\mu_i\rho} dz = -\sum_{i=1}^{N} k_i\mu_i \underbrace{\rho\, dz}_{dp/dz=-g\rho} \Rightarrow d\tau = \sum_{i=1}^{N} k_i\mu_i\frac{dp}{g}$$
$$= \sum_{i=1}^{N} k_i(p_{ref})\mu_i(p_{ref})\left(\frac{p}{p_{ref}}\right)^{a_i+b_i}\frac{dp}{g} \quad [60]$$

Here, k_i is the mass absorption coefficient for absorber i, which generally depends on pressure and temperature, while $\rho_i=\mu_i\rho$ is the corresponding density of the ith absorber, where μ_i is the mass mixing ratio (eqn [4]). In the second step, we substitute for the hydrostatic eqn [5]. In the last step, we use the fact that the profiles of μ_i and k_i are generally power laws of pressure:

$$\mu_i(p) = \mu_i(p_{ref})\left(\frac{p}{p_{ref}}\right)^{a_i}, \quad k_i(p) = k_i(p_{ref})\left(\frac{p}{p_{ref}}\right)^{b_i} \quad [61]$$

Water vapor is the main source of opacity in the Earth's atmosphere, whereas in other atmospheres, gases have similar pressure dependence of opacity and mass mixing ratio, such as N_2 and CH_4 on Titan. Furthermore, the sum over $k_i\mu_i$ in eqn [60] could be replaced with a single weighted-mean absorption coefficient k. Thus, we can integrate and normalize eqn [60] to get a power law dependence of optical depth on pressure, where the power is $n_p=a+b+1$,

$$dt \propto \left(\frac{p}{p_{ref}}\right)^{a+b} dp \Rightarrow \tau = \tau_0\left(\frac{p}{p_{ref}}\right)^{a+b+1}, \text{ or } \tau = \tau_0\left(\frac{p}{p_{ref}}\right)^{n_p} \quad [62]$$

Here, τ_0 is the gray IR optical depth at reference pressure p_{ref}, such as 1 bar. In tropospheres and lower stratospheres, often a well-mixed gas ($a=0$) with an absorption coefficient linearly dependent on pressure ($b=1$) occurs, such as H_2 in giant-planet atmospheres. In that case, $n_p=2$. The Earth's atmosphere is complicated, with a 8–12 μm window, a steep ($n_p=4$–5) τ–p scaling for the 6.3 μm water absorption band and water bands beyond about 20 μm, and an $n_p=2$ scaling for the 15 μm CO_2 band. But it turns out that $n_p=2$ is a good approximation for tropospheres and lower stratosphere for all the planets with thick atmospheres in the solar system, validated by comparison of numerical calculations (Robinson and Catling, 2012, 2014).

An $n_p=2$ power law in eqn [62] also has a physical basis. The primary absorption processes that cause IR opacity in thick tropospheres are pressure broadening and CIA (see **Section 10.13.3.5**). Because both processes rely on collisions, they tend to have an absorption coefficient $k_i \propto p$, so by integration of $d\tau \propto p$, it follows that $\tau \propto p^2$.

With an opacity–pressure power law scaling of eqn [62], we can rewrite our tropospheric convective profile with opacity as the vertical coordinate instead of pressure:

$$T = T_{ref}\left(\frac{\tau}{\tau_0}\right)^{\alpha_c(\gamma-1)/n_p\gamma} = T_{ref}\left(\frac{\tau}{\tau_0}\right)^{\beta/n_p} \quad [63]$$

This equation can be inserted into Schwarzschild's differential equation that describes the upwelling thermal radiative flux (eqn [44]) to give a differential equation in $dF^+/d\tau$, which can be solved (see the Appendix of Robinson and Catling (2012)) to give an analytic solution for the convective upwelling flux below the radiative–convective boundary in a gray troposphere:

$$F^+(\tau) = \underbrace{\sigma T_{ref}^4 e^{-D(\tau_0-\tau)}}_{\text{attenuated upwelling flux from } \tau_0 \text{ level}}$$
$$+ \underbrace{\frac{\sigma T_{ref}^4 e^{D\tau}}{(D\tau_0)^{4\beta/n_p}}\left(\Gamma_u\left(1+\frac{4\beta}{n_p}, D\tau\right) - \Gamma_u\left(1+\frac{4\beta}{n_p}, D\tau_0\right)\right)}_{\text{flux contribution from atmosphere above the reference level}} \quad [64]$$

Here, Γ_u is an upper incomplete gamma function, which is a mathematical function defined by

$$\Gamma_u(a,x) \equiv \int_x^\infty t^{a-1}e^{-t}dt \quad [65]$$

Although this looks complicated, from a practical perspective, the incomplete gamma function is evaluated as simply as a sine or cosine in computer programs, such as MATLAB, IDL, or Python. Thus, the thermal flux in eqn [64] can be instantly computed as a function of optical depth, if input parameters (n_p, β) and variables (T_0, τ_0) are specified.

In the preceding text, we have written analytic equations that govern radiative–convective equilibrium in planetary atmospheres with the gray approximation. The temperature follows a convective profile according to eqn [63] up to a radiative–convective boundary, where the temperature then obeys eqn [54] according to radiative equilibrium. There will be a tropopause temperature minimum above the radiative–convective boundary at an IR optical depth given by eqn [55] if the stratospheric shortwave absorption is large compared to IR opacity, that is, the condition of large k_{strato} in eqn [56]. **Figure 6(a)** is a schematic profile

that shows a purely radiative–convective equilibrium for an atmosphere without a stratospheric shortwave absorber. The temperature asymptotes to an isothermal stratosphere with a so-called skin temperature. The skin temperature is given by eqn [53] with $\tau = 0$, no downward thermal flux, and no solar atmospheric absorption. If we also neglect the internal heat flux, we get

$$\sigma T_{\text{skin}}{}^4 = \frac{F^{\odot}_{\text{tropo}} + F^{\odot}_{\text{strato}} + \cancel{F_i}}{2} \Rightarrow T_{\text{skin}}{}^4 = \frac{(\text{absorbed solar flux})}{2\sigma} = \frac{\sigma T_e{}^4}{2\sigma} \quad \Rightarrow T_{\text{skin}} = \frac{T_e}{2^{1/4}} \qquad [66]$$

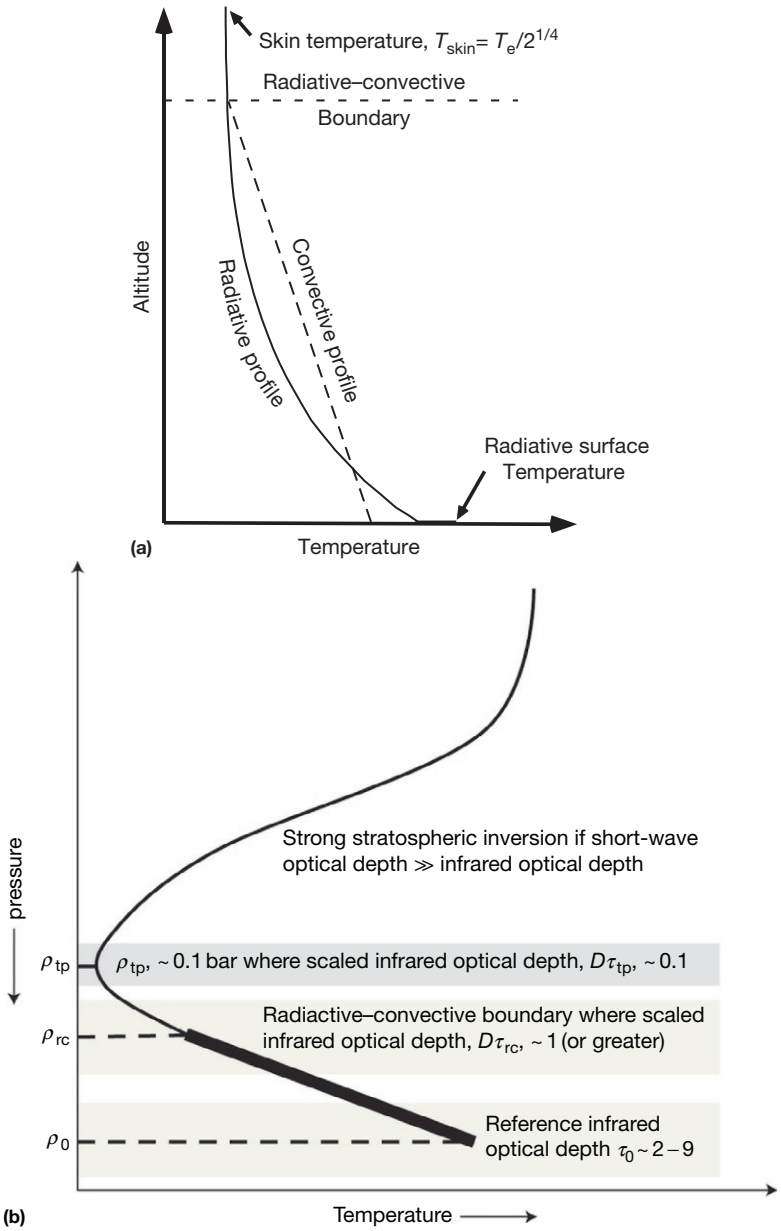


Figure 6 (a) A schematic profile of a radiative–convective equilibrium for an atmosphere without a shortwave absorber. A purely radiative profile (solid line) is unphysical. It is unstable to convection and has a discontinuity at the surface, so turbulence ensues and an adiabatic (i.e., convective) profile (dashed line) is established. The convective profile intersects the radiative profile at a level called a radiative–convective boundary with continuity of net flux and temperature. Without a stratospheric inversion, there is no tropopause minimum. In the absence of significant internal heat from the planet, the temperature asymptotes at altitude to a skin temperature that is a simple function of the effective temperature T_e, as shown. In our solar system, this schematic radiative–convective profile is roughly representative of the global mean profile of Venus. (b) A more general radiative–convective equilibrium profile where a strong shortwave stratospheric absorber is present. The conditions for the stratospheric inversion, tropopause minimum, and typical IR optical depths are shown (applicable for the Earth, Titan, and the giant planets). D is the diffusivity factor (see text). Reproduced from Robinson TD and Catling DC (2014) Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency. *Nature Geoscience* 7: 12–15.

where T_e is the effective temperature. Thus, the skin temperature of an Earth-like world without the properties that give it a stratospheric inversion would be $255\,\mathrm{K}/2^{1/4}$, which is 214 K. Although simple, the skin temperature actually gives a rough estimate for the tropopause minimum temperature; on Earth, the observed global mean tropopause temperature is 208 K (Han et al., 2011). **Figure 6(b)** shows a schematic for the more general case for an atmosphere with a stratospheric inversion.

Of course, gray radiation is an approximation, and numerical nongray calculations are required to replicate the details of atmospheres, for example, Goody and Yung (1989), p. 404–407; Thomas and Stamnes (1999), p. 450–453; and Satoh (2004), p. 380–383. Such models were first pioneered for the Earth by Manabe and coworkers (**Figure 7**) (Manabe and Strickler, 1964; Manabe and Wetherald, 1967). Since then, radiative and radiative–convective models have been applied to Venus (Bullock and Grinspoon, 2001; Crisp, 1986, 1989; Meadows and Crisp, 1996), Mars (Kasting, 1991; Marinova et al., 2005), Jupiter and Saturn (Appleby and Hogan, 1984), Uranus and Neptune (Appleby, 1986), Titan (McKay et al., 1991, 1999), and exoplanets (Heng et al., 2012; Robinson and Catling, 2012).

10.13.3.5 Absorption and Emission by Atmospheric Gases

In the previous subsection, we implicitly assumed that some gases absorb or scatter radiation at certain wavelengths (described as optically *or* radiatively active gases), whereas others do not (optically inactive). Why this is so involves an understanding of molecular spectroscopy.

In a spectrum from a planetary atmosphere, we see discrete lines, groups of lines called bands, windows of weak spectral absorption, and a continuum of smoother varying absorption. Our starting point recognizes that the total internal energy of a gas molecule can be written as

$$E_{\mathrm{total}} = E_{\mathrm{electronic}} + E_{\mathrm{vibrational}} + E_{\mathrm{rotational}} + E_{\mathrm{translational}} \qquad [67]$$

where $E_{\mathrm{electronic}}$ is the energy associated electron orbits and other terms represent vibrational energy, rotational KE, and translational KE, respectively. Molecular collisions depend on temperature through $E_{\mathrm{translational}}$ and redistribute energy between the various forms (if the air is dense enough) to establish local thermodynamic equilibrium where a local volume behaves like a blackbody. This was the assumption in the radiative transfer discussed in **Sections 10.13.3.3** and **10.13.3.4**.

A photon can be absorbed if its energy is close to a quantized energy jump within a molecule or it can be scattered by changing phase or direction. Transitions in rotational speed require less energy than vibrational ones, while electronic transitions require the most. Planck's law for photon energy, $E = h\nu = hc/\lambda$, indicates that types of transition correspond to general wavelength ranges (**Table 6**). However, absorption often occurs as combinations of transitions, such as rotational and vibrational.

An electric dipole moment due to separated positive charge and negative charge determines whether a molecule undergoes certain rotational or vibrational transitions. For example, the slight positive charge on hydrogen relative to oxygen in H_2O produces an electric dipole moment. The electric field in electromagnetic radiation torques the electric dipole and induces rotation. Isolated homoatomic molecules such as H_2 do not possess electric dipoles but dipoles can be induced by collisions. CH_4 and CO_2 also have no electric dipoles, but their bonds can bend, allowing a temporary electric dipole with associated rotational–vibrational transitions.

Some molecules possess a magnetic dipole moment. The magnetic field in an electromagnetic wave can torque a molecule with a magnetic dipole moment into rotation, analogous to a torqued needle on a compass. Many molecules do not have significant magnetic moments because electrons of opposite spin pair up. O_2 notably has two unpaired electrons and a magnetic dipole, which allows rotational bands in the microwave. In fact, satellite sensing of microwave emission from O_2 measures the Earth's atmospheric temperature over broad altitudes according to instrument weighting functions like those discussed in **Section 10.13.3.2** (Shine et al., 2008; Yang et al., 2013).

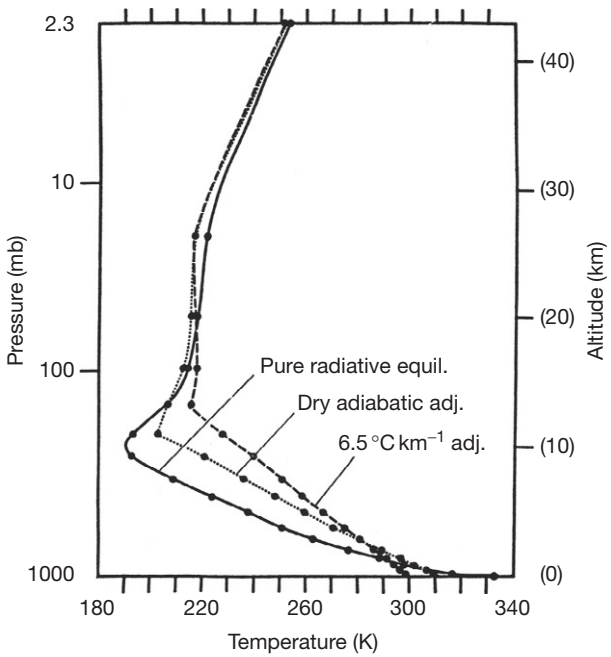


Figure 7 Results from radiative transfer modeling for the Earth that includes the effects of stratospheric ozone that causes a temperature inversion above a tropopause minimum. Three models are shown: purely radiative equilibrium, radiative–convective equilibrium with a dry adiabat, and radiative–convective equilibrium with a typical adiabat modified by water vapor. Reprinted from Manabe S and Strickler RF (1964) Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of Atmospheric Sciences* 21: 361–385, with permission from the American Meteorological Society.

Table 6 Typical wavelength ranges of molecular transitions

Dominant transition	*Wavelength range*	*Band*
Electronic	<1 μm	X-ray to UV to visible
Vibration	1–20 μm	Near-IR to far-IR
Rotation	>20 μm	Far-IR to microwave[a]

[a]H_2 can have rotational transitions at shorter wavelengths in warm, dense atmospheres (Wordsworth and Pierrehumbert, 2013).
IR, infrared; UV, ultraviolet.

For IR opacity and climate, the most important transitions are vibrational and rotational, so we discuss these next.

10.13.3.5.1 Rotational transitions

Rotating molecules have a rotational frequency that is quantized analogous to a household fan spinning at set speeds. A molecule's rotational KE is $^1/_2 I\omega^2$, where I is the moment of inertia. In general, there are three principal moments of inertia about three axes of symmetry through the center of mass of a molecule, denoted l_A, l_B, and l_C from smallest to largest. However, in many simple molecules, some moments of inertia are equivalent or zero. For example, the axis through the atoms of CO_2, a linear molecule (O=C=O), has $l_A \approx 0$, while the moments of inertia around two perpendicular axes are equal and can be designated as l. In quantum theory, the quantized total rotational angular momentum L, associated with a so-called linear rotor such as CO_2, is

$$L = \frac{h}{2\pi}\sqrt{J(J+1)} \qquad [68]$$

where $J = 0, 1, 2, \ldots$ is the rotational quantum number. The rotational KE for number J is

$$E_J = \frac{1}{2}I\omega^2 = \frac{(I\omega)^2}{2I} = \frac{L^2}{2I} = \frac{h^2}{8\pi^2 I}(J(J+1)) = hB(J(J+1)) \quad \text{where } B = \frac{h}{8\pi^2 I} = \text{rotational constant}$$

Only transitions between adjacent states J and $J+1$ are allowed, giving an energy change,

$$\Delta E = hB[(J+1)(J+2) - J(J+1)] = hB\left[J^2 + 3J + 2 - J^2 - J\right] = 2hB(J+1)$$

Thus, the photon frequency needed to cause a rotational transition in a linear rotor is

$$\nu = \frac{\Delta E}{h} = 2B(J+1) \qquad [69]$$

Consequently, a pure rotational absorption spectrum has a series of lines separated by frequency differences $\Delta\nu = 2B$. As a molecule rotates with increasing J, centrifugal force causes bonds to stretch, which requires a more complex treatment. Also, similar expressions can be derived for more molecules with many moments of inertia.

10.13.3.5.2 Vibrational transitions

Molecular bonds behave like springs, so a diatomic molecule has a classical simple harmonic resonant frequency ν_0. In quantum mechanics, however, the oscillation frequency and associated energy levels of a diatomic molecule are quantized as

$$\nu = \left(\mathrm{v} + \frac{1}{2}\right)\nu_0, \quad E_\mathrm{v} = h\nu_0\left(\nu + \frac{1}{2}\right)$$

where the letter 'vee,' $\mathrm{v} = 0, 1, 2, \ldots$, is a vibrational quantum number. The $\mathrm{v} = 0$ state has a finite zero point energy, $E_0 = h\nu_0/2$, and a transition from this state to the next highest, or vice versa, is the fundamental, which typically produces the strongest line because at atmospheric temperatures and pressures, most molecules are in the ground state. Large displacements of atoms in molecules corresponding to $\Delta\mathrm{v} = \pm 2, \pm 3, \ldots$ produce less intense overtones.

The frequency of combined vibrational and rotational transitions is slightly lower or higher than the pure vibration. Consequently, rotational–vibrational spectra have lines on either side of a pure vibrational frequency. The *P branch* is the low-frequency side corresponding to changes $\Delta J = -1$, while the *R branch* is the higher-frequency side with $\Delta J = +1$. The central frequency line with $\Delta J = 0$ is called the *Q branch*, which may not exist if quantum mechanics requires $\Delta J \neq 0$. Lines in the *P* and *R* branches are labeled according to the initial rotational quantum number J, that is, R(0), R(1),..., and P(0), P(1), etc. Superposed $\Delta J = -2$ or +2 electric quadrupole transitions can also occur. Instead of a PQR branch structure, there are the *O branch* ($\Delta J = -2$) and *S branch* ($\Delta J = +2$). Giant-planet atmospheres have electric quadrupole absorption lines of H_2 in the far-IR, for example.

Vibrational bands of CO_2 are important on Venus, Earth, and Mars. The bending of CO_2 produces a 15 μm rotation–vibration band – actually a PQR branch structure but individual lines are hidden by pressure broadening (see the succeeding text). The 4.3 μm asymmetrical stretch CO_2 band is important on Venus for absorbing a large amount of outgoing IR at depth because of Venus's high temperature. A vibration–rotation band of methane at 7.6 μm contributes IR opacity in the Earth's atmosphere but appears in emission in spectra of the giant planets and Titan due to warm CH_4-containing stratospheres.

10.13.3.5.3 Collision-induced absorption

Homoatomic molecules such as H_2 or N_2 lack permanent electric dipoles, but collisions at high-pressure distort the electronic structure and produce a temporary electric dipole, allowing vibrational and rotational transitions (Trafton, 1966; Trafton et al., 1998). CIA in various pairs of molecules creates significant IR continuum opacity in several planetary atmospheres: H_2–H_2 and H_2–He on the giant planets, N_2–N_2 and N_2–CH_4 in Titan's atmosphere (McKay et al., 1989), and CO_2–CO_2 on Venus (Moskalenko et al., 1979) or the early Earth (Kasting and Ackerman, 1986; Wordsworth et al., 2010). As first noted by Sagan (1977) H_2–N_2 and H_2–CO_2 CIA may have been important on prebiotic Earth (Wordsworth and Pierrehumbert, 2013) and early Mars (Ramirez et al., 2014). At Earth-like temperatures, H_2 is a strong greenhouse gas because its small moment of inertia causes widely spaced rotational energy levels (eqn [69]) that absorb across the thermal IR.

10.13.3.5.4 Broadening

Broadening processes are critical for planetary climates because otherwise, IR absorption lines would be delta-functions with negligible climatic effect. The three processes are the following:

- *Natural broadening*, in which a finite time for absorption or emission causes energy and frequencies to spread by Heisenberg's uncertainty principle. Apart from the UV, this broadening is tiny compared to the other two processes in the succeeding text.
- *Doppler broadening*, where random motion of molecules relative to the radiation source causes absorption and emission at frequencies that are Doppler-shifted relative to the

rest position of a line. This broadening tends to dominate in upper stratospheres and higher.

- *Pressure (or collisional) broadening*, where collisions add or remove energy during radiative transitions allowing emission or absorption over a broader range of photon energies. Pressure broadening is a key source of IR opacity in tropospheres (Section 10.13.3.4).

Pressure broadening produces a characteristic line shape called the *Lorenz profile* allowing absorption and emission far from the line center (Petty, 2006). A key property is that pressure-broadened lines have a width linearly proportional to pressure (i.e., the mass absorption coefficient relationship $k_\nu \propto p$, as assumed in Section 10.13.3.4 for tropospheres) unlike Doppler broadened lines, but both Doppler line broadening and pressure line broadening depend on temperature.

10.13.3.5.5 Continuum

A CIA continuum for giant planets and Titan was mentioned in the preceding text. For Earth, weak continuum absorption from 12.5 to 8 μm is mainly due to water vapor with absorption coefficient proportional to the square of the water vapor density. This absorption may be caused by the distant wings of very many lines in the far-IR or by dimers (H_2O•H_2O), trimers, and polymers of H_2O molecules, but the problem is not fully resolved. At shorter wavelengths, continuum absorption arises from well-known processes, including photoionization and photodissociation.

10.13.3.5.6 Numerical radiative transfer

Although gray models like those in Section 10.13.3.4 are indispensable for intuitive understanding, numerical radiative transfer calculations are needed for realism, which deal with hundreds of lines that often overlap. Line-by-line calculations explicitly sum the effects of all the contributions of individual lines. This accurate method is very computationally demanding for climate models. Consequently, two more practical approaches have been developed (Goody and Yung, 1989; Liou, 2002; Petty, 2006; Thomas and Stamnes, 1999; Vardavas and Taylor, 2007):

- *Band models* use an analytic form for the positions and strengths of lines. Over a spectral interval, lines can be represented by a mean line strength and spread in a random distribution or by some other spacing that is amenable to analytic formulation.
- The *correlated k-distribution method* is a statistical representation of the apparently haphazard variation of absorption coefficients k_ν over spectral interval $\Delta\nu$ with a smooth cumulative probability function of the occurrence of k_ν. Within $\Delta\nu$, the Planck function can be considered a function of temperature without frequency dependence. The order in which the k_ν's are summed then does not matter. In this way, integration can be done using bigger spectral intervals than in the detailed line-by-line spectrum.

Line-by-line calculations validate the aforementioned two methods and constrain the region of parameter space where the inherent assumptions are valid.

10.13.4 Photochemistry

10.13.4.1 Basic Principles of Chemical Kinetics

Clearly, through molecular properties like those discussed in the preceding text, atmospheric composition sets and controls planetary climates and atmospheric structures. But what sets composition? The concentration of any gas is determined by chemical kinetics, that is, a competition between source and sink fluxes, which depends on reaction rates. In general, atmospheres are never in chemical equilibrium because of the free energy from sunlight (or starlight for exoplanets), as well as interior or tidal energy fluxes. However, giant-planet atmospheres that convect air to pressurized, hot interiors, where reactions are efficient, are far closer to chemical equilibrium than shallower atmospheres on rocky planets such as the Earth or Mars. One can think of chemical equilibrium as telling you what 'should happen' if everything fully reacted, whereas kinetics determines what actually happens according to the speed of reactions. Thus, the starting point for atmospheric chemistry is to consider reaction rates. Then, we consider other principles before discussing specific planetary atmospheres.

A general two body reaction can be written as

$$bB + cC \rightarrow gG + hH \qquad [70]$$

The number of reactions per unit volume and unit time is given by the reaction rate,

$$R_r = k_r {n_B}^m {n_c}^n \qquad [71]$$

where n_i is number density (molecules cm^{-3}) of species i. Here, the exponents are related to the order of the reaction $m+n$ and are determined experimentally. A rate coefficient or rate constant k_r is established in the lab by mixing reactants and measuring reaction rates. The units of k_r depend on the order of the reaction and the concentration units. If $m=1$ and $n=0$, the reaction would be first order with k_r in units of s^{-1}; if $m=n=1$, the reaction would be second order with k_r in units of cm^3 $molecule^{-1}$ s^{-1}; and so on. The rates of change of species are related:

$$\frac{dn_B}{dt} = -b \times \text{rate} = -bR_r, \quad \frac{dn_G}{dt} = gR_r, \text{ etc., so}$$
$$-\frac{1}{b}\frac{dn_B}{dt} = -\frac{1}{c}\frac{dn_C}{dt} = \frac{1}{g}\frac{dn_G}{dt} = \frac{1}{h}\frac{dn_H}{dt} \qquad [72]$$

We can consider an example, oxidation of methane by a hydroxyl (OH) species:

$$OH + CH_4 \rightarrow CH_3 + H_2O \qquad [73]$$

The Earth's troposphere has concentrations $n_{OH} \sim 8 \times 10^5$ molecules cm^{-3} and $n_{CH_4} \sim 4.3 \times 10^{13}$ molecules cm^{-3}, while an average tropospheric rate constant for eqn [73] is $k_{rCH_4} = 3.6 \times 10^{-15} cm^3$ $molecule^{-1}$ s^{-1}. The units indicate a first-order reaction. By applying eqn [71], we can estimate the chemical lifetime of CH_4, that is, the average time a CH_4 molecule spends in the atmosphere before destruction:

$$\tau_{CH_4} = \frac{n_{CH_4}}{\text{destruction rate}} = \frac{n/_{CH_4}}{k_{rCH_4} n_{OH} n/_{CH_4}}$$
$$= \frac{1}{\left(3.6 \times 10^{s-15} cm^{-3} \text{ molecules} s^{-1}\right)\left(8 \times 10^5 cm^{-3}\right)}$$
$$\sim 12 \text{ years}$$

Thus, methane would be removed rapidly without resupply. Methane attains a nonnegligible abundance of ~1.8 ppmv in the Earth's present atmosphere because of a significant source flux: ~80–90% is from the Earth's microbial biosphere, and the rest is from geothermal sources (Kirschke et al., 2013). On Mars – with another oxidizing atmosphere – the lifetime of methane is ~300 years (Lefevre and Forget, 2009; Zahnle et al., 2011), so methane sources are very limited or nonexistent given in situ measurements of <1 ppbv CH_4 (Webster et al., 2013).

Generally, we deduce that a lifetime is related to chemical sinks as follows:

$$\begin{aligned} &X + Y_1 \xrightarrow{k_{r1}} \\ &X + Y_2 \xrightarrow{k_{r2}} \text{ products}, \quad \Rightarrow \text{chemical lifetime of X} = \tau_x = \frac{1}{\sum_i k_{ri} n_{Yi}} \\ &\text{etc.} \xrightarrow{k_{ri}} \end{aligned} \qquad [74]$$

Three-body (termolecular) reactions are also important in planetary atmospheres, for example, the formation of ozone (O_3) in the Earth's stratosphere from $O + O_2 +$ a third body. When two gases, A and B, combine, bond formation releases energy that must be removed to prevent the *adduct* (the form AB) from splitting. A third body 'M' takes away the energy, where 'M' represents any atmospheric gas, usually the most abundant one:

$$A + B + M \rightarrow AB + M \qquad [75]$$

Because a third body is needed, termolecular reactions are pressure-dependent up to a high-pressure limit, where the exact concentration of M no longer matters. Also, termolecular reactions are usually faster at lower temperatures because the molecules approach each other at slower speeds.

More generally, rate constants have an Arrhenius dependence on thermal energy kT, as

$$k_r = A\exp(-E_a/kT) \qquad [76]$$

where a pre-exponential factor A and activation energy E_a are determined experimentally. A few useful rules are the following: (i) Reactions between molecules and radicals (species with unpaired electrons, see Section 10.13.4.2) usually have positive E_a so that k_r increases with temperature. (ii) Fast reactions between two radicals generally have negative activation energy E_a so that k_r decreases with temperature. (iii) Reactions between nonradical molecules are normally very slow unless they react on particle surfaces. (iv) Favored reactions tend to be exothermic.

10.13.4.2 Why Free Radicals Are So Important

Appreciably fast atmospheric reactions, with few exceptions, involve at least one radical species. A free radical has an unpaired electron in its outer shell (e.g., atomic chlorine, Cl) and reacts strongly to obtain more stable bonding. An exception is ground-state atomic oxygen (O), a biradical that has eight electrons of which two are unpaired. Some literature indicates radicals with dots, for example, HO·.

Photochemistry is initiated by the Sun, loosely analogous to the spark plugs of an engine. Photons break up molecules (Section 10.13.4.3), creating radicals that transmit the free energy. Reactions of radicals with nonradical molecules make further radicals because electrons do not pair up, that is,

$$\begin{aligned} &\text{radical} + \text{nonradical} \rightarrow \text{radical} + \text{nonradical} \\ &\left(\text{e.g.,} \ \underbrace{\text{CN}}_{\text{nitrile radical}} + CH_4 \rightarrow \underbrace{CH_3}_{\text{methyl radical}} + \begin{array}{c}\text{HCN, formation of} \\ \text{hydrogen cyanide on Titan}\end{array} \right) \end{aligned} \qquad [77]$$

However, radical–radical reactions remove radicals in two ways: disproportionation into nonradicals

$$\begin{aligned} &\text{radical} + \text{radical} \rightarrow \text{nonradical} + \text{nonradical} \\ &\left(\text{e.g.,} \ \underbrace{C_2H_3}_{\text{vinyl (or ethenyl) radical}} + C_2H_3 \rightarrow \underbrace{C_2H_4}_{\text{ethylene}} + \underbrace{C_2H_2}_{\text{acetylene}}, \ \begin{array}{c}\text{a reaction at} \\ \sim 200\,\text{km} \\ \text{altitude on Titan}\end{array} \right) \end{aligned} \qquad [78]$$

and recombination with the formation of a single molecule

$$\begin{aligned} &\text{radical} + \text{radical} + M \rightarrow \text{nonradical} + M \\ &(\text{e.g., } NO_2 + OH + M \rightarrow HNO_3 + M, \text{ formation of nitric acid on Earth}) \end{aligned} \qquad [79]$$

By removing radicals, a chain of photochemical reactions initiated by sunlight is terminated. The net effect of all reactions is polymerization in reducing atmospheres, which can produce particles of hydrocarbons, sulfur, phosphorus, and so on, depending on the gas mixture, while in oxidizing atmospheres, the outcome is analogous to combustion, for example, full oxidation of CH_4 to water and CO_2 or oxidation of sulfur gases to sulfate aerosols (Hu et al., 2012, 2013).

Certain gas species can also be lost to planetary surfaces. 'Dry deposition' is interception with the surface and loss without the aid of precipitation, although the surface could be wet or dry. 'Wet deposition' is when raindrops, snow, or cloud or fog particles scavenge gases to the surface. For example, nitric acid formed in eqn [79] is dry-deposited onto the driest desert on Earth, the Atacama Desert (Catling et al., 2010; Michalski et al., 2004), whereas elsewhere, it usually dissolves in rain that falls to the surface. Deposition is how atmospheric chemistry can alter the geochemistry of a planetary surface. For example, the surface of Mars has accumulated sulfate salts that formed in its atmosphere from the oxidation of volcanic sulfur (Settle, 1979; Smith et al., 2014).

10.13.4.3 Photolysis

If species A is photolyzed, its number density n_A (molecules cm^{-3}) declines with a photodissociation rate,

$$A + h\nu \rightarrow \text{products}, \quad \text{rate} = \frac{\partial n_A}{\partial t} = -j_A n_A \qquad [80]$$

where j_A is the photolysis rate (or photodissociation) coefficient (s^{-1}). The probability Ψ_i per second for a molecule i to absorb radiation is related to j_i, as follows:

$$\Psi_i = \int_\lambda \sigma_{i\lambda} I_\lambda d\lambda \ \ s^{-1}, \quad j_i = \int_\lambda q_{i\lambda} \sigma_{i\lambda} I_\lambda d\lambda \ \ s^{-1} \qquad [81]$$

Here, $\sigma_{i\lambda}$ is the absorption cross section for species i (usually cm^2 $molecule^{-1}$), and $l_\lambda d\lambda$ is the photon flux crossing 1 cm^2 in spectral interval λ to $\lambda + d\lambda$. Absorption of a photon may sometimes lead to a different process (collisional deactivation, radiation, or breaking of a different bond), so $q_{i\lambda}$ is a dimensionless quantum yield for the photodissociation of interest.

10.13.4.4 Earth's Atmospheric Composition and Chemistry

With the principles in the preceding text, we can understand the Earth's atmosphere. The two key constituents responsible for its chemical character are very large amounts of O_2 and H_2O vapor compared with any other known atmosphere. The oxidizing chemistry caused by radicals associated with O_2 and H_2O is critical for radiative transfer because it renders the Earth's troposphere largely transparent to visible sunlight. At the same time, O_2 causes the creation of the ozone layer, shielding the Earth's surface from biologically harmful UV radiation with wavelengths less than about 300 nm.

In the stratosphere, the basic scheme for making O_3 from the O atoms of O_2 was deduced by Sydney Chapman in 1930 with the so-called Chapman reactions:

$$O_2 + h\nu(\lambda \leq 242\,\text{nm}) \rightarrow O + O \quad [82]$$

$$O + O_2 + M \rightarrow O_3 + M \quad [83]$$

$$O_3 + h\nu\,(\lambda \leq 1140\,\text{nm}) \rightarrow O + O \quad [84]$$

$$O_3 + O \rightarrow 2O_2 \quad [85]$$

Energy absorbed in ozone photolysis (eqn [84]) heats the stratosphere, causing its temperature inversion. O atoms derived from ozone photolysis cycle around to reform ozone via eqn [83]. But O_3 and O (the so-called odd oxygen species) return to stable O_2 molecules via eqn [85].

Chapman's scheme overpredicted the abundance of stratospheric O_3 compared to observations. In the 1970s, it was realized that minor constituents catalyze the removal of odd oxygen,

$$\begin{array}{l} X + O_3 \rightarrow XO + O_2 \\ \underline{XO + O \rightarrow X + O_2} \\ O_3 + O \rightarrow O_2 + O_2 \end{array} \quad [86]$$

where X is a radical: NO, OH, Cl, Br, or H.

In the 1980s, it was further shown that heterogeneous reactions (gas–solid) on particles in polar stratospheres could deplete the ozone, as opposed to the homogeneous reactions (gas–gas) reactions such as eqns [82]–[86]. In Antarctica, an ozone hole of low ozone column abundance appears each southern spring. Chlorine radicals released from the UV decomposition of anthropogenic CFCs are responsible (Russell et al., 1996). It was deduced that heterogeneous reactions convert unreactive forms of chlorine, such as HCl and chlorine nitrate into Cl_2 on polar stratospheric cloud particles:

$$\underbrace{ClONO_2}_{\text{chlorine nitrate}} + HCl \rightarrow Cl_2 + \underbrace{HNO_3}_{\text{nitric acid}} \quad [87]$$

This reaction sequesters odd nitrogen ($NO_x = NO + NO_2$, pronounced 'nox') into unreactive nitric acid, which prevents NO_x from turning reactive chlorine back into unreactive chlorine nitrate via

$$ClO + NO_2 + M \rightarrow ClONO_2 + M \quad [88]$$

Nitric oxide (NO) is naturally present because biogenically produced nitrous oxide (N_2O) reacts with an energetic oxygen atom (denoted $O(^1D)$) to generate two NO radicals.

In Antarctica, springtime sunlight releases Cl_2 from particles, which photolyzes into Cl radicals, causing ozone depletion via eqn [86]. The process is encouraged by a springtime stratospheric vortex of eastward winds that prevents inflow of new odd nitrogen or ozone from lower latitudes. Unlike Antarctica, buoyancy waves associated with north–south mountain ranges disrupt the development of a coherent Arctic vortex, so the Arctic lacks an equivalent ozone hole.

One aspect of the Earth's stratosphere that bears some commonality to other oxidizing atmospheres (e.g., Venus and early Mars) is that sulfur gases (sulfur dioxide, SO_2, and carbonyl sulfide, OCS) are oxidized into sulfate aerosols that make a thin haze called the Junge layer (at ~20–25 km altitude). The Junge layer scatters sunlight back to space and cools the Earth. The Junge layer is a photochemical haze of sulfates, analogous to Venus' clouds, although far thinner because the Earth's sulfur is mostly not gaseous but tied up in the form of sulfate and sulfide minerals on or below the Earth's surface. However, it is a general feature of oxidizing atmospheres with sulfur gases to produce sulfate hazes.

We now turn to the Earth's troposphere. All major gases (except Ar) in the Earth's atmosphere are mediated by biological cycles, including O_2, N_2, and CO_2. In addition, various trace gases have significant biological sources, such as CH_4 and bigger hydrocarbons, N_2O, OCS, dimethyl sulfide (CH_3SCH_3 and DMS), and carbon disulfide (CS_2).

A key factor for the Earth's tropospheric chemistry is the role of the dominant oxidizing radical, OH (Levy, 1971; Prinn, 2014). This species arises because of the presence of large amounts of water vapor and O_2, where the latter leads to the generation of tropospheric ozone (~10% of the ozone column abundance) in pathways involving NO_x. In the stratosphere, NO_x destroys ozone, but in the troposphere, NO_x makes ozone (Crutzen, 1979). To create ozone, O_2 and O need to be combined, so O atoms are required. These derive from NO_2, whose progenitor NO comes from lightning, soils, or transport from the stratosphere:

$$NMHC_S + NO + h\nu \rightarrow NO_2 + \text{products [a net reaction of many]}$$

$$NO_2 + h\nu(\lambda \leq 410\,\text{nm}) \rightarrow NO + O \quad [89]$$

NMHCs are nonmethane reactive hydrocarbons (C_2–C_5 compounds plus isoprene (C_5H_8), e.g., butane (C_4H_{10}) and ethylene (C_2H_4)), which have biogenic sources such as marine plankton and vegetation. The O atoms from eqn [89] react with O_2 to form ozone (eqn [83]). Oxidation of methane and carbon monoxide, catalyzed by NO_x, can also produce ozone. Then, ozone photolysis with relatively energetic photons generates exited O atoms that react with water vapor to generate OH radicals:

$$O_3 + h\nu(\lambda \leq 310\,\text{nm}) \rightarrow O_2 + O(^1D) \quad [90]$$

$$H_2O + O(^1D) \rightarrow OH + OH \quad [91]$$

OH is the 'detergent of the atmosphere' because it helps oxidize sulfur-containing gases (SO_2, H_2S, DMS, etc.) into sulfates, removes hydrocarbon gases such as CH_4 and NMHCs by initiating their oxidation, and converts NO_x into nitric acid. Without significant OH, the Earth's atmosphere would accumulate hydrocarbons and sulfurous gases and be very different.

10.13.4.5 CO_2 Atmospheres: Venus and Mars

Venus and Mars both have CO_2-dominated atmospheres and so have some common behavior (Krasnopolsky and Lefevre, 2013). For example, CO_2 does not significantly decompose into CO by photolysis on either planet because catalytic reactions recombine CO_2 from CO and O. Also, both planets have cold thermospheres caused by CO_2 radiative cooling, but Venus' thermosphere is colder, which is counterintuitive given Venus' location. Photochemistry lies behind this situation.

Of course, the recombination of CO_2 relies on radicals. On Mars, OH radicals (derived from water vapor or hydrogen peroxide photolysis) catalyze the recombination of CO and O in cycles (McElroy and Donahue, 1972; Parkinson and Hunten, 1972), for example,

$$\begin{aligned} &OH + CO \rightarrow CO_2 + H \\ &H + O_2 + M \rightarrow HO_2 + M \\ &HO_2 + O \rightarrow OH + O_2 \\ &\text{Net}: \; CO + O \rightarrow CO_2 \end{aligned}$$

Unlike Mars, where the chemistry is largely restricted to C–O–H, seven principal elements are involved on Venus: O–C–S–H–Cl–N–F. Chlorine atoms (derived from HCl photolysis) play an important role in CO_2 recombination on Venus, along with various catalysts:

$$\begin{aligned} &Cl + CO + M \rightarrow ClCO + M \\ &ClCO + O_2 + M \rightarrow ClCO_3 + M \\ &ClCO_3 + X \rightarrow Cl + CO_2 + XO \\ &XO + O \rightarrow X + O_2 \\ &\text{Net}: \; CO + O \rightarrow CO_2 \end{aligned}$$

where X = H, Cl, O, SO, or SO_2.

A fundamental difference of Venus compared with Mars (and Earth) is a total cloud cover. Venus' clouds of sulfuric acid are photochemical, originating from the photolysis of SO_2 and CO_2:

$$\begin{aligned} &SO_2 + h\nu \rightarrow SO + O \\ &CO_2 + h\nu \rightarrow CO + O \\ &SO_2 + O + M \rightarrow SO_3 + M \\ &SO_3 + H_2O \rightarrow H_2SO_4 \end{aligned} \qquad [92]$$

Unlike the Earth's clouds that require updrafts and cooling for condensation, photochemical haze occurs globally on Venus, analogous to Titan's haze. Sulfur on Venus also disproportionates, that is, some gets oxidized, while the rest is reduced. Molecules of sulfur atoms and sulfane (sulfur–hydrogen) absorb visible and UV radiation and may be responsible for cloud coloring.

Venus' upper thermosphere is colder than Mar's (~250 K vs. ~270 K) because O atoms are ~10 times more abundant. There is a greater concentration of O atoms on Venus versus Mars because of more photodissociation of CO_2 and a smaller-scale height (eqn [7]) for the same temperature, given the dependence on gravitational acceleration (Krasnopolsky, 2011). On both planets, chemistry produces vibrationally excited CO_2 (denoted by an asterisk),

$$O + CO_2 \rightarrow CO_2^* + O \qquad [93]$$

which decays and radiates 15 μm photons through an optically thin path to space, causing cooling.

10.13.4.6 Reducing Atmospheres: Giant Planets and Titan

Much could be said about the photochemistry on Titan and the giant planets (e.g., Yung and Demore, 1999), but here, I focus on one general aspect – the fate of hydrogenated gases. Although Titan and the giant planets have reducing atmospheres, hydrogen escapes to space on Titan but not from the giant planets where H_2 is the major constituent. This changes the fate of carbon species on Titan compared to the giant planets and is also relevant for the processing of nitrogen, phosphorus, and sulfur species on the giant planets.

Methane is the most abundant carbon gas in the cold reducing atmospheres of the solar system (although CO is thermodynamically favored in hot Jupiters). Unlike the rapid conversion to CO_2 and water vapor that occurs in oxidizing atmospheres (**Section 10.13.4.1**), methane is destroyed when UV photons with wavelength of <145 nm break C–H bonds. The radicals that are released combine to make bigger hydrocarbons in the following overall schematic net reaction:

$$\begin{aligned} (CH_4 \text{ molecules}) + h\nu &\rightarrow \text{radicals} + \text{hydrogen} \\ &\rightarrow (\text{large hydrocarbons}) + \text{hydrogen} \end{aligned} \qquad [94]$$

On Titan, for example, the lifetime of atmospheric methane against complete photolysis loss is 30–100 My (Yung et al., 1984). The hydrocarbons can be as simple as ethane (C_2H_6) or as complex as polyaromatic hydrocarbons, which are molecules consisting of linked benzene rings. On Titan, organic molecules also incorporate nitrogen to form nitriles (organics with C≡N groups) (Atreya et al., 2009). But on Titan, hydrogen escapes, so the organics accumulate on the surface as solids or liquids, and methane is destroyed (**Figure 8(a)**). However, on the giant planets, hydrogen does not escape and is the main gas. There, hydrocarbons eventually get mixed down to deep, hot (~1000 K) layers where they are hydrogenated and converted back into methane.

Similar cycles occur on the giant planets with nitrogen, phosphorus, and sulfur species. Ammonia (NH_3), phosphine (PH_3), and hydrogen sulfide (H_2S) are photolyzed to make hydrazine, elemental phosphorus, and sulfur-containing solids. However, sedimentation and mixing to depth cyclically reconstitute ammonia, phosphine, and hydrogen sulfide (**Figure 8(b)**).

10.13.5 Atmospheric Dynamics

10.13.5.1 Equations of Motion

Atmospheric dynamics concerns the study of atmospheric motions and their relationship to heating and forces (Holton, 2004; Houghton, 2002; Mak, 2011; Vallis, 2006). Winds are

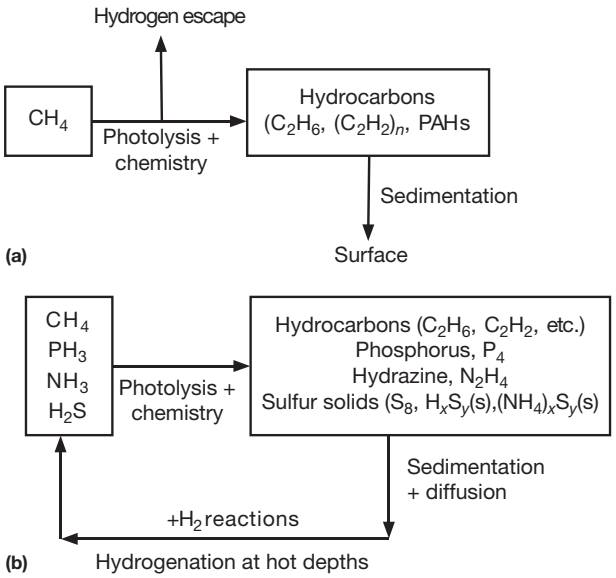


Figure 8 (a) Net hydrocarbon production on Titan, associated with the escape of hydrogen. PAHs are polyaromatic hydrocarbons. (b) The closed cycles of hydrogen-bearing species on giant planets. (Drawn by the author.)

critical for climate because they transport heat (e.g., from tropics to cold poles) and affect the formation of clouds by carrying condensables. Winds also redistribute non-condensables (e.g., O_3 in the Earth's stratosphere), modify surfaces (e.g., dunes on Mars or Titan), and can drive secondary circulations (e.g., ocean currents on Earth).

Atmospheric motions are usually considered from the perspective of an observer corotating with a planet. A rotating frame is accelerating, and in accelerating frames, apparent forces arise in the opposite direction to real accelerations. For example, scales in an upwardly accelerating elevator would incorrectly measure weight without accounting for a downward apparent force. In a rotating frame, centrifugal and Coriolis apparent forces arise. Centrifugal forces are known from everyday experience, but the Coriolis force is less familiar and, consequently, often illustrated by a ball thrown from the center of a merry-go-round to a catcher at its edge. If the merry-go-round does not rotate, the ball reaches the catcher in a straight line, but with counterclockwise rotation, the catcher rotates around and the ball misses him; the ball appears to be deflected to the right of its motion from the viewpoint of the corotating catcher and thrower. The analogy is imperfect because the ball also experiences a radially outward centrifugal force. But the basic picture is analogous to looking down upon a planet spinning west to east: the Coriolis force acts to the right of motion (e.g., a wind vector) such as in the Earth's northern hemisphere. Viewed from below, the merry-go-round's rotation is clockwise and the Coriolis force acts to the left of the motion, akin to the effect on winds in the Earth's southern hemisphere.

Newton's second law for motion on a rotating planet (e.g., see Taylor, 2005) is

$$\underbrace{\mathbf{a}}_{\text{acceleration on co-rotating frame}} + \underbrace{(2\boldsymbol{\Omega}\times\mathbf{v})}_{\text{coriolis acceleration}} + \underbrace{[\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r})]}_{\text{centripetal acceleration}} = \underbrace{\frac{\mathbf{F}_g}{m}}_{\text{force of gravity}} + \underbrace{\frac{\mathbf{F}}{m}}_{\text{all other forces}} \qquad [95]$$

where $\boldsymbol{\Omega}$ is the planet's angular velocity vector, the acceleration and velocity are measured by a corotating observer, and the last term includes pressure gradient forces, frictional forces, and external tidal forces caused by gravitational interaction with the Sun or parent star, from a moon, or from a planet being orbited if we are considering the atmosphere of a large moon. Usually, the centripetal acceleration is taken to the right-hand side, treated as 'centrifugal force per kg' and, combined with gravity to form an effective, measured gravity g. A modified g vector is valid because rotating planets deform into oblate surfaces as a result of the centrifugal force pulling outward at the equator. Simplifying eqn [95] by neglecting tidal torques, we get the usual basic form of the momentum equation in atmospheric dynamics,

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \underbrace{\frac{-1}{\rho}\nabla p}_{\text{preesure gradient force/kg}} - \underbrace{(2\boldsymbol{\Omega}\times\mathbf{v})}_{\text{Coriolis force/kg}} - g\mathbf{k} + \mathbf{F}_{\text{vise}} \qquad [96]$$

where $\mathbf{F}_{\text{visc}}$ is the frictional or viscous force per unit mass, $\mathbf{v}$ is the wind vector, and $\mathbf{k}$ is an upward unit vector. It is possible that tidal forces should not be neglected for long-term climate models rather than weather models. But surprisingly little work has been done to examine this issue (Houben, 2013).

Equation [96] can be expanded into local Cartesian (x, y, z) coordinates on a sphere with equations in three velocity components (e.g., see Holton (2004), Chapter 2): a zonal wind u in the west to east or 'zonal' x direction, a meridional wind v in the south-to-north y direction, and a vertical wind w upward in the z direction, that is, $\mathbf{v} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$. The resulting momentum equations simplify for large-scale motions, if we ignore friction and vertical motion (given that vertical winds are usually much smaller than horizontal ones) (e.g., see Andrews (2010) or Holton (2004)):

$$\text{Vertical (hydrostatic equation)}: \quad \frac{\partial p}{\partial z} = -g\rho \qquad [97]$$

$$\text{Zonal momentum}: \quad \frac{\mathrm{d}u}{\mathrm{d}t} - \underbrace{\left(f + \frac{u\tan\phi}{a}\right)}_{\text{Coroilis term}\quad\text{Centrifugal term}} \mathrm{v} = -\frac{1}{\rho}\frac{\partial p}{\partial \mathrm{x}} \qquad [98]$$

$$\text{Meridional momentum}: \quad \frac{\mathrm{d}v}{\mathrm{d}t} + \underbrace{\left(f + \frac{u\tan\phi}{a}\right)}_{\text{Coroilis term}\quad\text{Centrifugal term}} u = -\frac{1}{\rho}\frac{\partial p}{\partial y} \qquad [99]$$

Here, a is the planetary radius at the local x, y, z origin, the latitude is ϕ, and f is the Coriolis parameter defined by

$$f = 2\Omega\sin\phi \qquad [100]$$

where Ω is the planet's angular velocity. Specifically, $\Omega = 2\pi/\tau_{\mathrm{d}}$ where τ_{d} is the rotation period (=a sidereal day), for example,

$\Omega = 2\pi/(86\,164\text{ s}) = 7.29\times10^{-5}$ rad s^{-1} for Earth. If put on the right-hand side, terms f_v and $-f_u$ are x and y components of the Coriolis force per kg that act to the right of the wind in Earth's northern hemisphere. Terms $-(u\tan\phi/a)v$ and $+(u\tan\phi/a)u$ are x and y components of cyclostrophic acceleration, which are important on slowly rotating planets (see Section 10.13.5.2).

In eqns [98] and [99], the derivatives in v and u are substantial *or* material derivatives. The dX/dt form is the (Lagrangian) rate of change of flow-field quantity X with respect to time following a fluid parcel, unlike dX/dt, the (Eulerian) rate of change at a fixed point. They are related by

$$\underbrace{\frac{dX}{dt}}_{\text{substantial derivative}} \equiv \underbrace{\frac{\partial X}{\partial t}}_{\text{local derivative}} \underbrace{+u\frac{\partial X}{\partial x}+v\frac{\partial X}{\partial y}+w\frac{\partial X}{\partial z}}_{\text{advective terms}} = \frac{\partial X}{\partial t}+\mathbf{v}\cdot\nabla X \qquad [101]$$

Here, $\mathbf{v}$ is the velocity vector and we remind ourselves that $\mathbf{a}\cdot\mathbf{b} = a_1b_1+a_2b_2+a_3b_3$, while the gradient operator is $\nabla = \partial/\partial x + \partial/\partial y + \partial/\partial z$. For example, if 2-D spatial contours of temperature T are fixed in time ($\partial T/\partial t = 0$) and an observer moves eastward with velocity u in time. Δt between 250 and 270 K contours, then eqn [101] gives $dT/dt = u\nabla T = (\Delta x/\Delta t)\times(\Delta T/\Delta x) = 20\text{ K}/\Delta t$. Alternatively, if the 250 K contour moves with the wind speed and replaces the 270 K downstream contour, $dT/dt = 0$ because the local change $\partial T/\partial t$ ($-20\text{ K}/\Delta t$) is balanced by advection of temperature, $u\nabla T$. 'Advection' (Latin 'carry to') is the transport of an atmospheric property by the velocity field. The difference between advection and convection is that advection usually refers to the predominantly horizontal, large-scale atmospheric motions, while convection describes the mostly vertical, locally induced motions.

10.13.5.2 Balance Regimes: Geostrophic, Cyclostrophic, and Thermodynamic

10.13.5.2.1 Geostrophic balance

On planets that rotate fairly rapidly, such as Earth, Mars, and the giant planets, the Coriolis force is important. In the Earth's northern hemisphere (NH), the Coriolis force acting to the right of large-scale flow is frequently approximately balanced by low pressure to the left and high pressure to the right, that is, a pressure gradient force. Thus, on NH weather maps, winds blow counterclockwise around a 'low' cyclonic system or clockwise around a 'high' anticyclonic system. The balance of Coriolis and pressure gradient forces is called geostrophic balance. Because $f \gg u/\tan\phi$, eqns [98] and [99] provide this balance as

$$u_g = -\frac{1}{f\rho}\left(\frac{\partial p}{\partial y}\right)_z, \quad v_g = \frac{1}{f\rho}\left(\frac{\partial p}{\partial x}\right)_z \qquad [102]$$

The z subscript indicates a partial derivative evaluated along an altitude surface, while the g subscript denotes geostrophic winds. In deriving eqn [102], we assumed $d(u \text{ or } v)/dt \ll f$ to eliminate d/dt terms. This condition is true if the dimensionless Rossby number (Ro), which is the ratio of the Lagrangian acceleration to Coriolis acceleration, is <0.1 or so, where

$$\text{Rossby number, } Ro = \frac{\text{Relative acceleration}}{\text{Coriolis acceleration}} = \frac{U^2/L_h}{fU} = \frac{U}{fL_h} \qquad [103]$$

Equation [103] uses a scale analysis. The variables U and L_h are characteristic wind speeds and horizontal length scales, respectively. For example, Jupiter has $U \approx 100\text{ m s}^{-1}$, $L_h \approx 100\,000\text{ km} = 10^8$ m, and $f \approx 1.8\times10^{-4}\text{ s}^{-1}$ at 45°N. Thus, $Ro \approx 0.01$, so rotation dominates and geostrophy applies. Similarly, $Ro \ll 1$ for other giant planets, Earth, and Mars. But at the equator, $\phi = 0$ and $f \to 0$, so Ro cannot be small. The result is that circulations on rotating planets are more symmetrical with latitude close to the equator but rotation affects flows strongly at higher latitudes, producing big eddies.

Geostrophy predicts that zonal winds should typically increase in speed with altitude. Usually, the distribution of sunlight creates a temperature decrease from equator to pole, so columns of air on average shrink toward the pole because they are cooler, which must create an increasing meridional pressure gradient at higher altitudes. Since $\partial p/\partial y$ grows with altitude, the zonal wind u_g will grow too by eqn [102]. This link between horizontal temperature gradients and the wind shear – that is, change of wind speed and direction with height – is captured in the thermal wind equation, which is derived from eqn [102] by substituting for $\rho = p/\bar{R}T$ from the ideal gas law (eqn [2]), differentiating w.r.t. z and then substituting the hydrostatic eqn [6]:

$$f\frac{\partial v_g}{\partial z} \approx \bar{R}T\frac{\partial}{\partial x}\underbrace{\left(\frac{1}{p}\left(\frac{\partial p}{\partial z}\right)\right)}_{-g/\bar{R}T} = \frac{-g\cancel{\bar{R}}T}{\cancel{\bar{R}}}\frac{\partial}{\partial x}\left(\frac{1}{T}\right) = \cancel{f}\left(\frac{g}{T\cancel{f}}\frac{\partial T}{\partial x}\right)$$
$$= \frac{g}{T}\frac{\partial T}{\partial x}$$

An approximation sign is used for ignoring vertical variations in T. With analogous manipulation of the zonal geostrophic wind equation, the thermal wind equations in u_g and v_g are

$$f\frac{\partial u_g}{\partial z} \approx -\frac{g}{T}\frac{\partial T}{\partial y}, \quad f\frac{\partial v_g}{\partial z} \approx \frac{g}{T}\frac{\partial T}{\partial x} \qquad [104]$$

These equations (or slightly higher approximations) allow calculation of winds on other planets (e.g., Smith et al., 2001). After deriving temperatures from radiances measured by an orbital instrument, eqn [104] implies the geostrophic winds. Often, no surface wind is assumed as a boundary condition at the expense of some accuracy.

10.13.5.2.2 Cyclostrophic balance

The analysis in the preceding text applies to rapidly rotating planets, but for slowly rotating planets, the balance is different. For example, on Venus, $L_h \approx 1000\text{ km} = 10^6$ m, $U \approx 10\text{ m s}^{-1}$, and $f = 2\Omega\sin\phi = 2(2.98\times10^{-7}\text{s}^{-1})(\sin\ 45) = 4.2\times10^{-7}\text{s}^{-1}$ at 45°N, so the $Ro \approx 10$ from eqn [103] and geostrophy is invalid. If the atmosphere itself rotates sufficiently rapidly over some altitude range, the substantial derivatives and f terms are negligible in eqns [98] and [99], and a cyclostrophic balance and a cyclostrophic thermal wind apply, as follows:

$$\text{(a)}\ \underbrace{\frac{u^2\tan\phi}{a} = -\frac{1}{\rho}\frac{\partial p}{\partial y}}_{\text{cyclostrophic balance}};$$

$$\text{(b)}\ \underbrace{\frac{\tan\phi}{a}\frac{\partial(u^2)}{\partial z} \approx -\bar{R}T\frac{\partial}{\partial y}\left(\frac{1}{p}\frac{\partial p}{\partial z}\right) = -\frac{\bar{R}\cancel{T}}{\cancel{\bar{R}}T\cancel{f}}\frac{g}{\cancel{f}}\frac{\partial T}{\partial y} = -\frac{g}{T}\frac{\partial T}{\partial y}}_{\text{cyclostrophic thermal wind equation}} \qquad [105]$$

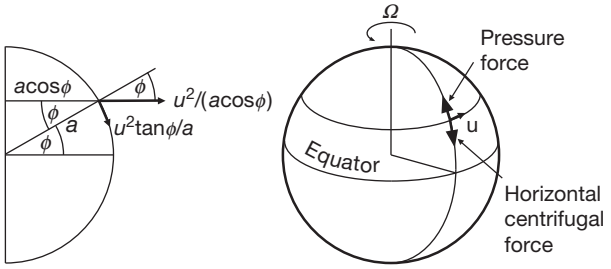


Figure 9 The geometry of cyclostrophic balance and the balance of forces on a slowly rotating planet such as Venus or Titan. Here, ϕ is latitude, a is planetary radius, $\mathbf{u}$ is the zonal wind vector, and Ω is a planetary rotation vector. (Drawn by author.)

Equation [105](a) is a balance between centrifugal force per unit mass and an equator-to-pole pressure gradient force (Figure 9). At latitude ϕ, the centrifugal acceleration due to the zonal wind is $u^2/(a\cos\phi)$, because $a\cos\phi$ is the distance to the rotation axis. On Venus and Titan, wind speeds increase with altitude consistent with a cyclostrophic thermal wind according to Pioneer Venus (Schubert, 1983; Seiff et al., 1980) and Huygens probe (Bird et al., 2005) measurements, respectively. Equation [105] should also apply to slowly rotating exoplanets.

10.13.5.2.3 Thermodynamic balance

Large-scale ascent and descent produce expansion and compression, which can change temperatures from those expected with simple radiative–convective equilibrium (Section 10.13.3.4). The physics derives from the first law of thermodynamics, written $dq = c_p dT - \alpha dp$, where $\alpha = 1/\rho$, from eqn [12]. Dividing by δt and c_p and taking the limit $\delta t \to 0$,

$$\underbrace{\frac{dT}{dt}}_{\text{time variable component}} + \underbrace{w\left(\frac{\partial T}{\partial z} + \frac{g}{c_p}\right)}_{\text{workdone}} \underbrace{-\frac{\alpha}{c_p}\left(\frac{dp}{dt}\right)}_{\approx -g\rho w} = \underbrace{\frac{Q}{c_p}}_{\text{heat gain}} \Rightarrow \underbrace{w\left(\frac{\partial T}{\partial z} + \frac{g}{c_p}\right)}_{\text{static stability}} \approx \underbrace{\frac{Q}{c_p}}_{\text{heat gain}} \quad [106]$$

Here, $Q = dq/dt$ [J kg^{-1} s^{-1}] is the diabatic heating per kg caused by radiation, convection, or conduction. In the middle part of eqn [106], we used eqn [101] to define T' as the horizontal (x, y) and time variable component of temperature and T as a horizontal and time average temperature. Since the term in T' is usually small, we neglect it for the final expression, which is a thermodynamic energy equation. This describes a rate of temperature change (K s^{-1} or K day^{-1}) for a parcel of air with vertical velocity and diabatic heating. We can replace the diabatic heating term with a Newtonian cooling approximation, in which temperature relaxes back to an equilibrium reference temperature field described by T_{eqm} as $Q/c_p = -\alpha_N(T - T_{eq})$, where $\alpha_N = 1/\tau_{rad}$ where τ_{rad} is a radiative time constant. Recognizing static stability (eqn [17]) in eqn [106], we can write thermodynamic balance as

$$w(\Gamma_a - \Gamma) = -\alpha_N(T - T_{eqm}) \quad [107]$$

This equation indicates that heating or cooling from vertical motion is balanced by radiative relaxation back to an equilibrium temperature. The radiative time constant can be estimated as

$$\tau_{rad} = \frac{\text{heat change per m}^2}{\text{absorbed flux per m}^2} = \left(\frac{p_e c_p(\Delta T_e)}{g}\right) \Big/ \left(\frac{(1-A)S_p}{4}\right) = \frac{4p_e c_p(\Delta T_e)}{(1-A)S_p g} \quad [108]$$

where p_e is the pressure at the emission level, g is the gravitational acceleration, ΔT_e is the temperature change forced by the absorption of solar radiation (~2 K for Earth or 80 K for Mars), c_p is the specific heat capacity, A is the Bond albedo, and S_p is the top-of-atmosphere solar flux. This uses eqns [8] and [23].

We can see the effect of eqn [107] in latitude-averaged temperature structures of planetary atmospheres. Static stability $(\Gamma_a - \Gamma)$ is generally positive, so if we measure $T < T_{eqm}$, then w is positive and large-scale ascent produces temperatures colder than equilibrium from cooling by expansion. Conversely, if $T > T_{eqm}$, large-scale descent produces compressional warming. In Mars' equinox atmosphere (Figure 10), one's naive expectation of warmer tropics and colder poles is reversed at pressures below ~0.5 mbar. Large-scale Hadley circulation cells (discussed in the succeeding text) produce descent and compressional warming at high latitudes and ascent and cooling at low latitudes.

10.13.5.3 Zonal Mean Meridional Circulation: Hadley Cells and Thermally Driven Jets

Jet streams or jets are significant features of the circulations of planetary atmospheres, defined as zonal wind maxima, whether eastward or westward. The style of atmospheric circulations is closely connected with jet stream structure, which varies greatly between planets. Jets can be single or multiple in each hemisphere. They may be strong or weak, meandering in latitude or straight. They are generally associated with transient eddies or waves whose zonal wave numbers (the number of complete waves around a latitude circle) vary between planets from wave number one on Venus to wave number >40 on Jupiter.

It is convenient to think of two types of jets: those driven directly by the equator-to-pole temperature gradient (thermally driven jets) and those driven indirectly by eddies (eddy-driven jets), respectively (Lee and Kim, 2003). In large-scale atmospheric dynamics, an eddy means a departure from a steady and uniform east–west flow, which results in variations in pressure with longitude. We consider thermally driven jets first.

The subtropical jet that occurs in the Earth's atmosphere at 30–40° latitude is thermally driven and associated with the descending branch of a large overturning of the tropical atmosphere called the Hadley circulation. This circulation is a large loop of airflow where air warms and rises in the tropics, cools and sinks at higher latitudes, and returns to the tropics near the surface to complete the loop, all under the influence of planetary rotation (Figure 11). George Hadley (1685–1786) invented the concept to explain tropical trade winds (Hadley, 1735). Equatorial-moving air is deflected by the Coriolis force toward the west and produces persistent westward tropical winds. In the subtropics, sinking causes high-pressure zones of dry air that also explain deserts such as the Sahara.

Hadley imagined a circulation symmetrical about the equator, but we know now that the Earth's Hadley circulation is

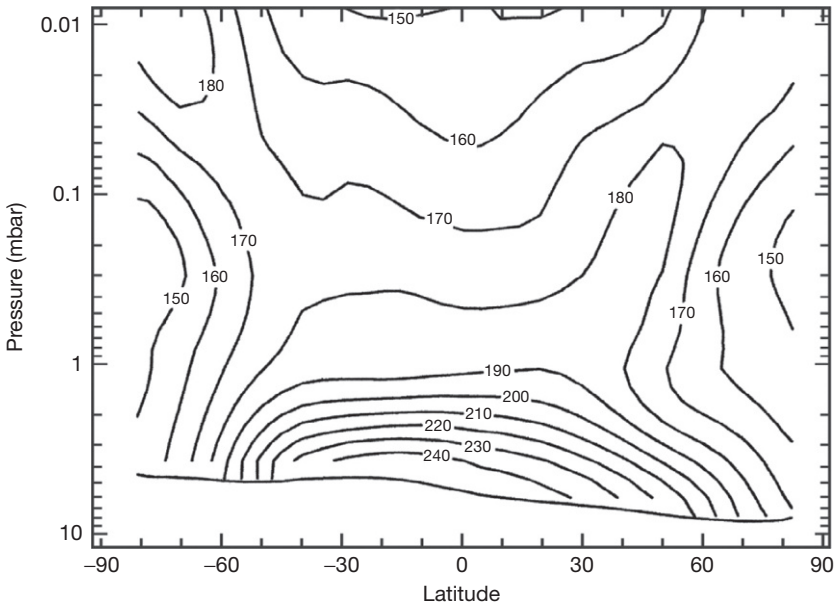


Figure 10 Contours of the zonal mean meridional temperatures (in K) at northern hemisphere autumnal equinox on Mars derived from measurements by the Thermal Emission Spectrometer (TES) instrument on NASA's Mars Global Surveyor orbiter. The solid line at the lower boundary of the temperature contours is the zonal mean atmospheric pressure at the surface. Adapted from Smith MD, Pearl JC, Conrath BJ, and Christensen PR (2001) Thermal Emission Spectrometer results: Mars atmospheric thermal structure and aerosol distribution. *Journal of Geophysical Research* 106: 23929–23945 with permission.

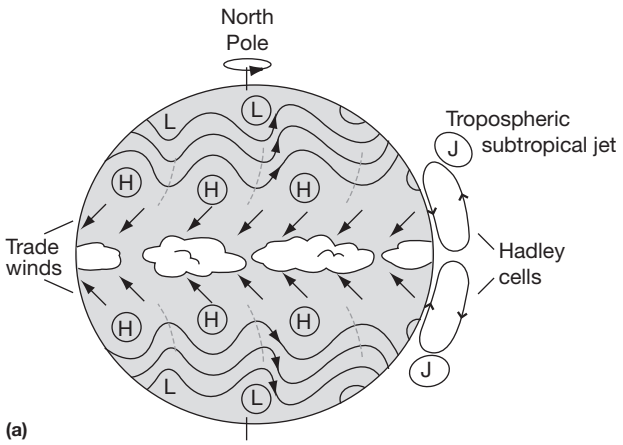


Figure 11 Idealized symmetrical Hadley cells on Earth. Moist air rises near the equator tropics causing a belt of clouds. The air aloft travels poleward, acquiring a fast zonal speed by conservation of angular momentum. This results in a subtropical jet near the descending branch of the Hadley cell. At the surface, air returns to complete the loop, causing westward trade winds in the tropics. In each hemisphere, a different circulation is found beyond the maximum latitudinal extend of the Hadley circulation. In midlatitudes, there are waves that are found in the eastward winds. (Adapted from Wallace JM and Hobbs PV (2006) *Atmospheric Science: An Introductory Survey*, 2nd edn. Burlington, MA: Elsevier Academic Press, International geophysics series. Used with permission.)

symmetrical in neither latitude nor longitude (Cook, 2004; Dima and Wallace, 2003; Lindzen and Hou, 1988). Rising air occurs as relatively small-scale convection associated with heavy rainfall in the intertropical convergence zone (ITCZ) where low-level southwestward trade winds from the northern hemisphere converge with northwestward trade winds from the south. The ITCZ is generally located somewhat off the equator (Figure 12). The descending branches of the Hadley circulation are also not uniform in longitude. Subsidence takes place preferentially in high-pressure areas over the subtropical Atlantic, Pacific, and Indian Oceans.

Monsoons are associated with the Hadley circulation and exist on Earth and Mars. A monsoon is a seasonal reversal of winds associated with latitudinal movement of the ITCZ, and the monsoon behaves as an intensification of the Hadley circulation in particular longitude ranges (e.g., Schneider (2006)). On Earth, the ITCZ is offset from the equator because of planetary obliquity and the distribution of oceans and continents. Surface winds curve across the equator and turn from westward to eastward due to the Coriolis force and become a monsoon circulation. The Martian orbit has a relatively high eccentricity with more incoming solar radiation near the southern summer winter solstice, and there is asymmetrical topography between hemispheres, causing the Hadley circulation to be stronger than during southern winter solstice (Haberle et al., 1993b; Richardson and Wilson, 2002). The lower Hadley branch can be channeled by large-scale topography to form monsoon-like circulations (Joshi et al., 1994, 1995).

A key physical concept behind the Hadley circulation is conservation of angular momentum. In the simplest model, one assumes that angular momentum per unit mass is conserved in the upper branch of the Hadley circulation. At the equator, assuming no zonal wind, the angular momentum per unit mass is Ωa^2, where a is the planetary radius. If angular momentum per unit mass is conserved in transport of air parcels from the equator to latitude ϕ, a zonal wind u is acquired, as follows:

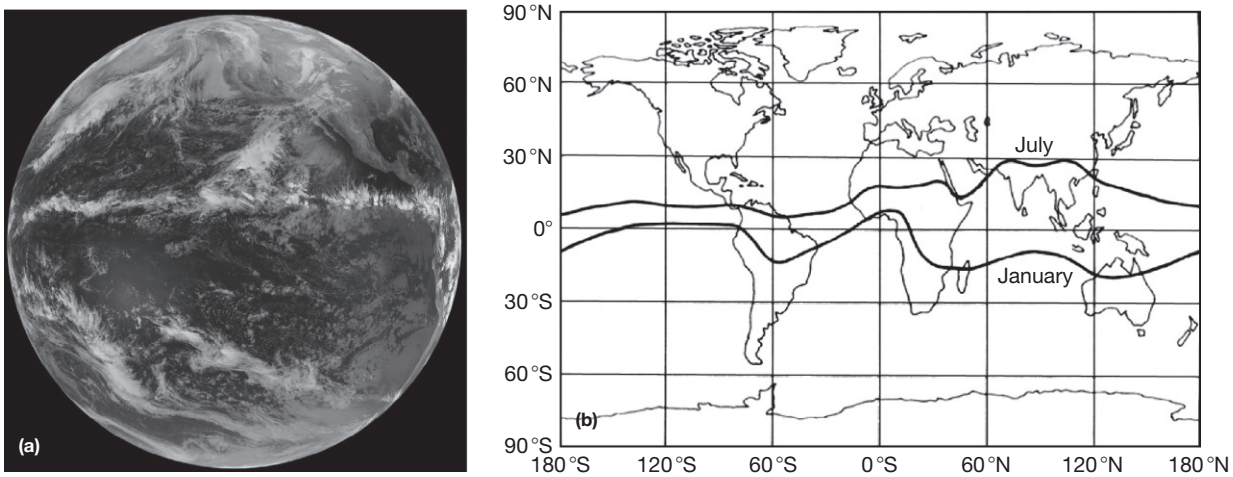


Figure 12 (a) The intertropical convergence zone (ITCZ) of the trade winds is visible on Earth as a band of cloud across the eastern Pacific from west to east (Courtesy: NOAA/NASA GOES project). (b) The latitude of the ITCZ changes seasonally and the annual average latitude of the ITCZ is north of the equator unlike an idealized symmetrical Hadley circulation.

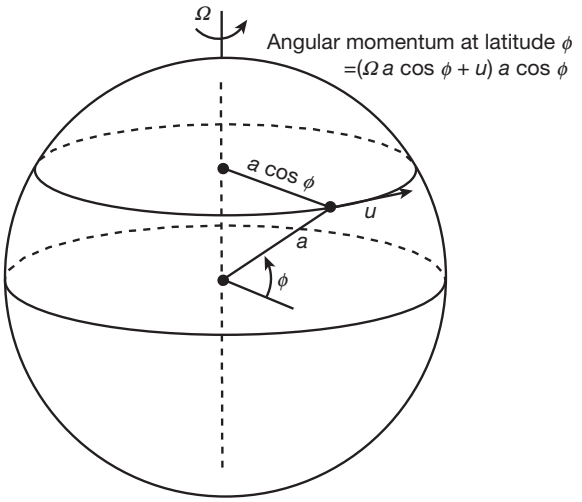


Figure 13 Zonal angular momentum acquired at latitude ϕ in poleward motion of air in the upper branch of the Hadley circulation helps in understanding the Hadley circulation. Rapid zonal wind speeds u are produced, which is responsible for a subtropical jet at the most poleward latitude to which the Hadley circulation extends. (Drawn by the author.)

$$\underbrace{\Omega a^2}_{\text{ang.momentum@equator}} = \underbrace{(\Omega a\cos\phi + u)a\cos\phi}_{\text{angular momentum@latitude }\phi} \Rightarrow u\cos\phi$$

$$= \underbrace{\Omega a\left(1-\cos^2\phi\right)}_{\sin^2\phi} \Rightarrow u = (\Omega a)\frac{\sin^2\phi}{\cos\phi} \qquad [109]$$

A point on the planet's surface at latitude ϕ moves with tangential velocity $\Omega a \cos\phi$, to which we add the zonal wind speed, u (Figure 13). The radius to the planetary rotation axis is $a\cos\phi$, so that angular momentum conservation is given by the first equality, and we solve for u. Imagine air in the upper branch of the Hadley circulation carried from the equator, where $u=0$, to latitude 30°. Equation [109] predicts $u = 134$ m s^{-1} at 30° Thus, we expect strong subtropical eastward winds at altitude.

The subtropical jet core in the upper branch of the Earth's Hadley circulation seasonally reaches ~30–70 m s^{-1} – nowhere as high as predicted in the preceding text because eddies from the midlatitudes provide westward angular momentum that decelerates the upper branch of the Hadley cell (Schneider, 2006). Of course, friction causes slow surface winds, so that winds must increase with altitude near the jet. We can look at the thermal wind eqn [104] in 'reverse': a large wind shear must have a correspondingly big meridional temperature gradient $\partial T/\partial y$ in the vicinity of the jet, unlike the smooth variation of equilibrium temperature with latitude that we might expect. Large-scale ascent is associated with $T < T_{eqm}$ on the equatorial side of the jet. Conversely, large-scale descent occurs on the polar side where $T > T_{eqm}$.

If one assumes that heating and cooling balance across a Hadley cell, then it is possible to derive an equation for the maximum latitude ϕ_H for how far a Hadley cell extends, which varies as $\phi_H \propto g^{0.5}/(\Omega a)$ (Held and Hou, 1980). However, such a theory underpredicts the meridional mass transport (kg s^{-1}) of the Hadley circulation by an order of magnitude. Other models suggest that midlatitude eddies confine the Hadley circulation. Possibly, the Hadley circulation winds and temperatures become unstable, which prevents penetration into midlatitudes. Energy carried from the tropics is transferred from the Hadley cells to eddies and waves characteristic of midlatitudes. Using such ideas, Held (2000) derived a weaker dependence on rotation of $\phi_H \propto (g^{0.5}/(\Omega a))^{0.5}$. Earth-like general circulation model simulations confirm that the Hadley circulation contracts in latitude range with increasing rotation rate (Del Genio and Suozzo, 1987; Navarra and Boccaletti, 2002; Showman et al., 2013). With slow rotation, if the

$Ro \gtrsim 10$, a planet behaves as an 'all tropics' world because Hadley cells extend to high latitudes (Showman et al., 2013).

As slowly rotating bodies, Venus and Titan should have Hadley circulations extending to high latitude, but relatively little evidence proves this. On Venus, ~20 W m^{-2} of sunlight reaches the surface (Titov et al., 2007), so the Hadley circulation must be weak. Surface wind markings from Magellan radar have an equatorward direction at higher latitudes and an east–west direction nearer the equator consistent with a large-scale Hadley circulation (Greeley et al., 1997). Venus absorbs 100 W m^{-2} near its cloud tops (~60–70 km), which should drive a second Hadley-like circulation above the clouds (Baker and Leovy, 1987). A strong temperature gradient and weak maximum in the zonal wind speed at ~55° latitude in both hemispheres indicates this circulation's probable extent. On Titan, low-latitude dunes indicate eastward flow, opposite to the expected trade winds (Jaumann et al., 2009). Possibly, the dunes form during periods when equatorial zonal winds reverse and strengthen (Tokano, 2009, 2010) (see **Chapter 10.12**).

10.13.5.4 Eddy-Driven Jet Streams

In terrestrial midlatitudes, we find an eddy-driven polar jet stream associated with systematically structured eddies superimposed on the zonal mean flow. A familiar example of such eddies are traveling cyclones and anticyclones. There is also a stationary wave structure due to the distribution of oceans and continents. On the giant planets, multiple jets are probably eddy-driven also (Ingersoll, 1990; Williams, 1978) (**Figure 14**). On Mars and Venus, the single, primarily thermally driven jet in each hemisphere is also altered by large-scale stationary and transient eddies.

To understand eddy-driven jets, we introduce concepts of rotation in fluid dynamics. Rather than quantify spin in terms of velocity and distance to a center of rotation, vorticity is used, a single quantity that describes the spin of a fluid element at a point, equivalent to the curl of the wind vector, curl $\mathbf{v} = \nabla \times \mathbf{v}$. In 2-D flows, ignoring the velocity dependence on altitude and the vertical velocity, relative vorticity ζ is a measure of spin about a vertical axis on a planet's surface:

$$\zeta = \nabla_h \times \mathbf{v}_h = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) \times (u, v) \quad \Rightarrow \zeta = -\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \qquad [110]$$

The sign of ζ follows a right-hand rule: with fingers in the direction of curved flow, 'thumb up' is positive. Vorticity measures both change of velocity transverse to the flow (horizontal shear) and curvature of the flow. In solid-body rotation, relative vorticity is 2ω (where ω is angular velocity): a shear contribution of the change in tangential velocity with radius, $\partial V/\partial r = \partial(\omega\ \mathrm{r})/\partial r = \omega$, adds to curvature from the change in the tangential velocity with angle, $V(\partial(\text{angle})/\partial(\text{path})) = V(2\pi/2\pi r) = V/r = \omega$. Cyclonic relative vorticity is always in the same direction as the component of the planet's rotation, and a cyclonic vortex is always a low-pressure zone. In the Earth's northern hemisphere, cyclonic flow is counterclockwise but would be clockwise in the southern hemisphere. Anticyclonic relative vorticity is in the opposite direction to the component of planetary rotation.

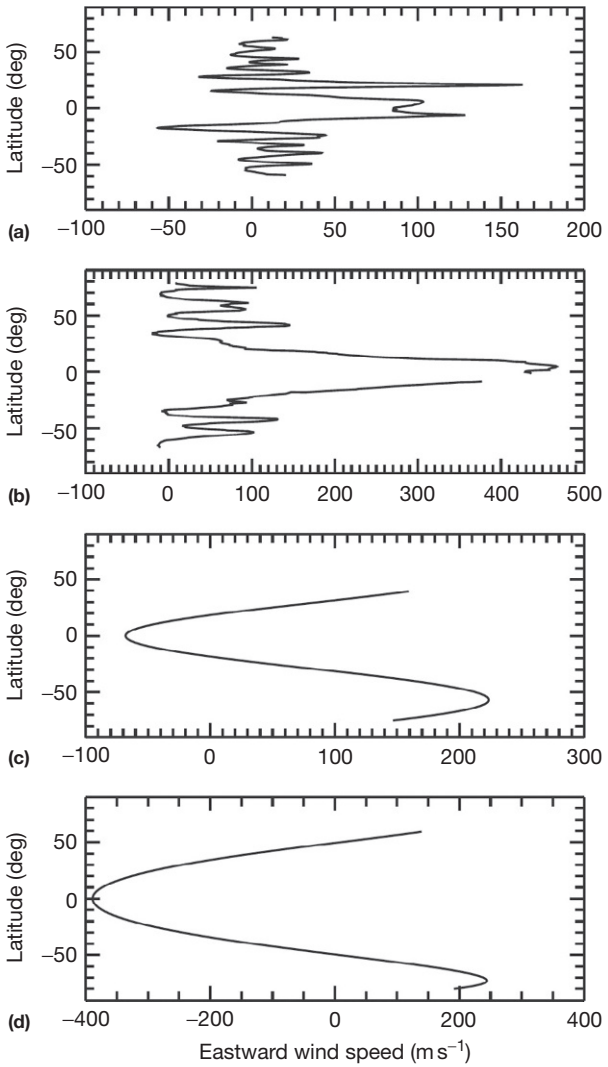


Figure 14 Jet streams on the giant planets are shown in zonal mean zonal winds from cloud tracking on (a) Jupiter, (b) Saturn, (c) Uranus, and (d) Neptune. The peaks, whether eastward or westward, correspond to jets. Jupiter and Saturn have 20 east–west jets, with a strong prograde jet over the equator, while Uranus and Neptune appear to have 3 broad jets, with strong retrograde jet over the equator. Adapted from Showman AP, Cho JYK, and Menou K (2010) Atmospheric circulation of extrasolar planets. In: Seager S (ed.) *Exoplanets*, pp. 471–516. Tucson, AZ: University of Arizona Press.

An element of fluid has a component of the planet's spin as well as any relative vorticity. The solid planet has a vorticity of 2Ω, and its vertical component of $2\Omega\sin\phi$ at latitude ϕ is the planetary vorticity, equivalent to the Coriolis parameter, f (eqn [100]). If we add relative vorticity (ζ) and planetary vorticity (f), we obtain the absolute vorticity η:

$$\eta = \zeta + f = \left(-\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) + f \qquad [111]$$

In adiabatic, inviscid flow in a thin homogeneous layer absolute vorticity, η, is conserved, that is, stays constant with time, if there is no friction or stirring. We denote the latter by a vorticity forcing term S_η:

$$(a)\ \text{no friction or stirring}\ \frac{d\eta}{dt}=0,$$

$$(b)\ \text{with friction or stirring}\ \frac{d\eta}{dt}=S_\eta \qquad [112]$$

Here, the substantial derivative is $d/dt=\partial/\partial t + u\partial/\partial x + v\partial/\partial y$.

Equation [112](b) can be used to understand how jets are forced by eddies. A common technique in fluid dynamics is linearization in which variables are represented by mean values and perturbations and inserted into equations; then certain terms are eliminated to obtain equations that can be solved analytically. Consider a uniform zonal mean flow and an eddy fluctuation deviation that is primed, $u=\overline{u}+u'$, along with a small meridional perturbation $v=v'$. Primed variables represent parts that vary in longitude, associated with eddies. The average of the product of fluctuation components, for example, $\overline{u'v'}$, is the covariance and is nonzero only when u' and v' are correlated, that is, in phase. Thus, applying a perturbation to eqn [112](b), we have

$$\left(\frac{\partial}{\partial t}+(\overline{u}+u')\frac{\partial}{\partial x}+v'\frac{\partial}{\partial y}\right)(\overline{\eta}+\eta')=S_\eta'$$

$$\Rightarrow\left(\frac{\partial}{\partial t}+\overline{u}\frac{\partial}{\partial x}\right)\eta'+v'\frac{\partial}{\partial y}\overline{\eta}=S_\eta' \qquad [113]$$

Here, $\overline{\eta}$ has no dependence on t or x, and we eliminate small terms. If we multiply eqn [113] by η', take the zonal average, and delete terms with a zonal mean of zero, we get an

$$\underbrace{\overline{\eta'\frac{\partial\eta'}{\partial t}}}_{\frac{1}{2}\frac{\partial(\eta')^2}{\partial t}}+\cancel{\overline{u}\overline{\eta'\frac{\partial\eta'}{\partial x}}}+\overline{v'\eta'}\frac{\partial\overline{\eta}}{\partial y}=\overline{S_\eta'\eta'}\Rightarrow \underbrace{\overline{v'\eta'}}_{\text{poleward vorticity transport by eddies}}$$

$$=-\underbrace{\frac{\partial\overline{(\eta')^2}/\partial t}{2\,\partial\overline{\eta}/\partial y}}_{\text{vorticity variance term}}+\underbrace{\frac{\overline{S_\eta'\eta'}}{\partial\overline{\eta}/\partial y}}_{\text{eddy forcing term}} \qquad [114]$$

Examine eqn [114]. Because f increases with latitude and $\partial f/\partial y$ is positive, the meridional gradient of absolute vorticity $\partial\overline{\eta}/\partial y$ is generally >0. So, the generation of vorticity variance by eddy stirring is associated with poleward flux of vorticity across some latitude circle ϕ_1 by the eddies represented by $\overline{v'\eta'}$ at latitude ϕ_1. This increases vorticity in the whole area of the planet poleward of ϕ_1.

It turns out that an increase of vorticity above latitude ϕ_1 can drive a prograde jet at ϕ_1 (here, the word prograde means in the same direction as the rotation of the planet, e.g., eastward for Earth; retrograde means the opposite direction). The vorticity (rotation at a point) is related to the circulation (rotation over an area) around a closed curve enclosing that area by Stokes' theorem. Writing down this theorem and applying it to the whole domical region (a spherical cap) north of latitude ϕ_1, we have

$$\underbrace{\Gamma_C}_{\text{circulation }[m^2s^{-1}]}=\oint\underbrace{V_s}_{\text{velocity along arc ds}}ds=\iint_{\text{area A}}\eta dA$$

$$\Rightarrow 2\pi(a\cos\phi_1)\overline{u}$$

$$=\int_{\lambda=0}^{2\pi}\int_{\phi=\phi_1}^{\pi/2}\eta a^2\cos\phi d\phi d\lambda$$

$$=2\pi[\eta]a^2\int_{\phi=\phi_1}^{\pi/2}\cos\phi d\phi \qquad [115]$$

The result means the following: a poleward vorticity flux by eddies that increases the average vorticity $[\eta]$ in the domical region poleward of ϕ_1 also increases the mean zonal wind $\overline{u}$ at ϕ_1.

Conversely, if eddies drift poleward to another latitude ϕ_2 where there is dissipation, then $\overline{S_\eta'\eta'}$ is negative. At latitude ϕ_2, the eddy vorticity flux will be equatorward (negative), and a retrograde jet will be forced via eqn [115]. Thus, if eddies propagate north or south, prograde and retrograde jets can be generated and multiple jets could form on rapidly rotating planets.

Another way to look at eddy-driven jets is to consider the north–south flux of momentum from the eddies. The northward eddy momentum across a latitude circle is $\rho v'$ (kg m^2 s^{-1}) so that the poleward flux of eastward eddy momentum is $\int\overline{u'(\rho v')}dA$ where the overbar is a zonal average and dA is an element of area of an imaginary great wall around the latitude circle. If the northward eddy wind and eastward eddy wind are correlated, $\overline{u'v'}>0$, and there is northward flux of zonal momentum per unit mass. Negative $\overline{u'v'}$ corresponds to a southward flux. From vector addition of the primed velocities, eddy wind component covariance corresponds to a tendency for tilted eddies. Where $\overline{u'v'}>0$, there is SW to NE tilt of the eddies. Conversely, $\overline{u'v'}<0$ corresponds to systematic NW to SE tilt. So, air parcels traveling along a SW–NE trajectory and then back feed momentum to an eastward jet to the north. More generally, if the shear of the zonal wind along a meridian, $d\overline{u}/dy$, and $\overline{u'v'}$ have the same sign, eddies accelerate jets. Eddies are said to transport momentum upgradient (Held, 1975; Thompson, 1971). This appears to be true on Jupiter (Salyk et al., 2006). On Earth, we find a departure from the symmetrical sinusoid form of waves and a systematic eastward tilt in eddies in midlatitudes so that southward motions have a smaller eastward component than northward motions (Starr, 1948), which causes net angular momentum transport (Figure 15).

10.13.5.5 Planetary Waves and Barotropic and Baroclinic Instabilities

Jets meander in latitude because of waves in the circulation. The most important class of waves in this regard are Rossby waves, which are oscillations that arise primarily in the horizontal because conservation of absolute vorticity acts to restore flows displaced meridionally to the latitude of origin. Consider an air parcel moving eastward with zero relative vorticity ($\zeta=0$) that is perturbed northward, perhaps because of flow over topography. Planetary vorticity f increases northward

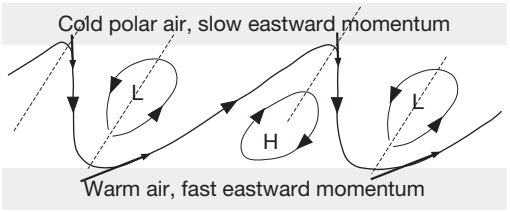


Figure 15 Tilts in the eastward winds cause momentum to be transported from the equator to poles. (These tilts were also drawn on Figure 11 as dashed lines). (Drawn by the author.)

(eqn [100]), so relative vorticity ζ must become negative to keep absolute vorticity constant ($\eta = \zeta + f$). Remembering the right-hand grip rule, negative ζ is anticyclonic curvature (clockwise in the Earth's northern hemisphere), which sends the parcel southward, back to its initial latitude (**Figure 16**). Then, an overshoot beyond the initial latitude will induce cyclonic curvature and send the parcel back northward. The air parcel sinusoidally oscillates as it moves downstream making a Rossby wave. Meanders in the Earth's midlatitude jet stream with typical wave numbers of 4 to 6 are guided by Rossby waves. Low-pressure cyclones tend to form in the troughs of the waves, and high-pressure anticyclones tend to form in the crests. Similar Rossby waves have been observed on Mars in midlatitudes in spring and winter with wave numbers 1–2 (Hollingsworth and Barnes, 1996; Read and Lewis, 2003; Wilson et al., 2002).

The restoring effect due to the gradient in planetary vorticity (f) is called the beta effect where beta $= \beta = \partial f/\partial y$ is the meridional gradient of planetary vorticity. If we assume a zonal mean zonal wind and differentiate absolute vorticity η (eqn [111]) w.r.t. y, we get

$$\frac{\partial \eta}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2}, \quad \text{where } \beta \equiv \frac{\partial f}{\partial y} = \underbrace{\frac{\partial (2\Omega \sin\phi)}{\partial y}}_{\delta y = a\delta\phi,\ \text{spherical geometry}}$$

$$= \frac{1}{a}\frac{\mathrm{d}f}{\mathrm{d}\phi} = \frac{2\Omega\cos\phi}{a} \qquad [116]$$

Usually, β is much larger than $-\partial^2\overline{u}/\partial y^2$ so $\partial\eta/\partial y \approx \beta$ provides the restoring effect.

It is possible to derive sinusoidal functions in x, y, and t that represent Rossby waves. The procedure is to rewrite the vorticity eqn [113] in terms of a single variable, a stream function, ψ, which is a scalar field for a fluid flow with contour lines that are streamlines of the flow, defined as $u = -\partial\psi/\partial y$ and $v = -\partial\psi/\partial x$ (e.g., see Holton (2004), pp. 213–217). The Rossby wave solutions in ψ are found to have a dispersion relation, meaning how the eastward phase speed c of the Rossby wave varies with wavelength, as follows:

$$c = \frac{\omega}{k} = \overline{u} - \frac{\beta}{(k^2 + n^2)}$$

$$= \text{phase speed relative to planet's surface} \qquad [117]$$

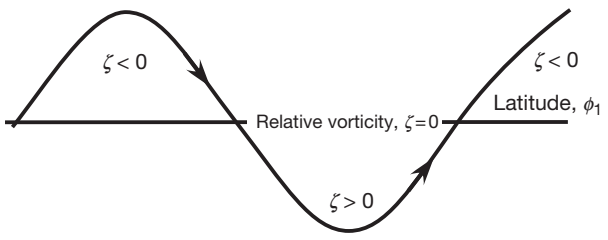


Figure 16 Illustration of the mechanism of Rossby waves in the *x*-*y* plane: a restoring effect caused by the gradient of absolute vorticity. A fluid parcel in the northern hemisphere moves eastward through the wave pattern. As it crosses latitude ϕ_1, the relative vorticity $\zeta = 0$. North of ϕ_1, conservation of absolute vorticity requires $\zeta < 0$, so the trajectory curves to the south anticyclonically. When it shoots over latitude ϕ_1, to conserve absolute vorticity $\zeta > 0$ and the trajectory curves northward cyclonically. (Drawn by the author.)

Here, $k = 2\pi/\lambda_x$ and $n = 2\pi/\lambda_y$ are zonal and meridional wave numbers, where λ_x and λ_y are zonal and meridional wavelengths, respectively. In the simple case where $n = 0$, the waves propagate eastward relative to the planet and westward relative to the zonal wind, at phase speeds that increase with increasing wavelength or smaller k. Setting $n = 0$ and $c = 0$ in eqn [117] gives $k^2 = \beta/\overline{u}$ for stationary waves with a west–east wavelength of $\lambda_x = 2\pi(\overline{u}/\beta)^{0.5}$. For example, in the Earth's middle troposphere at 45°N, $\beta \sim 1.6 \times 10^{-11}\ \mathrm{m^{-1}\,s^{-1}}$, the stationary wavelength is ~7000 km if $\overline{u} = 20\,\mathrm{m\,s^{-1}}$, and four such wavelengths would encircle the Earth.

Strictly horizontal flows with no temperature variations (parallel isobars and isotherms) are called barotropic and such flows are stable when $\beta - \partial^2\overline{u}/\partial y^2 0$. However, if the second term – the curvature of the zonal wind profile with latitude – becomes bigger than β, then instability is produced called the barotropic instability or Rayleigh–Kuo instability. Then, growing unstable eddies derive their energy from the KE of the zonal mean flow. In prograde jets, the zonal wind decreases to the right of the flow where $\partial^2\overline{u}/\partial y^2$ can be large. Barotropic instability produces unstable eddies that propagate to latitudes where they decay and accelerate retrograde zonal mean flow. Numerical simulations of jets forced by the eddies indicate a meridional length scale associated with this instability, characteristic of the width of the jets, called the Rhines radius, $r_{\mathrm{Rh}} \sim (2U/\beta)^{1/2}$, where U is the typical speed of the jet (Rhines, 1975). On giant planets, vortices stop growing at this size in the meridional directional and stretch out east–west, around a circle of latitude. Thus, the number of bands on the face of a gas giant is roughly a semicircle meridian divided by the Rhines radius, $-(\pi a)/r_{\mathrm{Rh}}$ (Menou et al., 2003).

In real atmospheres, we must consider vertical variations. General flows with vertical wind variations, nonzero vertical velocity, and horizontal temperature gradients (thus, mutually inclined isobars and isotherms) are called 'baroclinic.' Instead of vorticity, an analogous quantity, potential vorticity (PV), is introduced to allow for the vertical stretching or shrinking of fluid columns. Shrinking of the radius of a column of spinning fluid increases the spin rate, like an ice skater drawing in their arms. PV is conserved for a fluid element in the absence of mixing with other air parcels or frictional loss of KE and, in a constant density fluid, can be defined as

$$P_{\mathrm{v}} = \frac{\eta}{h} = \frac{(f + \zeta)}{h}\left(\mathrm{m^{-1}s^{-1}}\right) \qquad [118]$$

where h is the height of a short segment of a fluid column (e.g., Wallace and Hobbs, 2006, p. 289). Here, the word potential crops up again. In potential temperature (**Section 10.13.2.4**), we had a quantity remaining the same in adiabatic expansion, although the actual temperature changed. Potential vorticity is conserved for a fluid element, but the element's relative vorticity changes with changing latitude or column height. When density ρ varies with height, a more general version of PV is Ertel's potential vorticity Q_{P}:

$$Q_{\mathrm{P}} = \frac{\eta}{\rho}\frac{\partial\theta}{\partial z} \qquad [119]$$

where θ is potential temperature and the sign of $\partial\theta/\partial z$ indicates static stability (**Section 10.13.2.4**). In adiabatic flow, if static stability $\partial\theta/\partial z$ increases, shrinking and horizontal expansion

of fluid columns causes the absolute vorticity n to decrease according to eqn [119], if Q_P stays constant. The converse also applies. Since Q_P involves potential temperature, it can be generated and dissipated by diabatic heating, stirring, or friction.

Ertel's PV can be calculated from temperature and pressure fields, but analytic approaches often use quasigeostrophic (QG) potential vorticity q_g, which is an approximate form of PV valid when the *Ro* is small and flows are approximately geostrophic. Equations based around QGPV are helpful in studying large-scale motions characteristic of atmospheric circulations. An equation analogous to eqn [113] can be derived but with absolute vorticity η replaced by QGPV, q_g, which again can be written in terms of a single stream function and solved to give 3-D planetary wave solutions (e.g., see Andrews, 2010). The key difference is that such waves have an amplitude perturbation ψ' that depends on height. Analogous to $^1/_2(\text{mass}) \times (\text{velocity})^2$, the wave KE density $\sim \rho\psi'^2$ is independent of height in the absence of dissipation. So, for this quantity to remain constant, the amplitude grows as $\psi' \sim \rho_0^{-0.5}$, where ρ_0 is the background density. Such a relationship is used in photochemical models of planetary atmospheres where turbulence (at least to certain altitudes) is often parameterized proportional to $\rho^{-0.5}$. The same $\rho^{-0.5}$ dependence also applies to the vertical amplitude growth of buoyancy waves (**Section 10.13.5.7**).

The meridional gradient of zonal mean QGPV is somewhat analogous to eqn [116] but has an extra term that depends on the vertical shear of the zonal mean zonal wind (e.g., see Salby (2012), p. 465, or Andrews (2010), p. 128 for derivation of QGPV equations):

$$\frac{\partial \overline{q}_g}{\partial y} = \beta - \frac{\partial^2 \overline{u}}{\partial y^2} + \left(\frac{1}{\rho_0}\right)\frac{\partial}{\partial z}\left[\left(\rho_0 \frac{f_0^2}{N_B^2}\right)\frac{\partial \overline{u}}{\partial z}\right] \qquad [120]$$

Here, f_0 is the value of the Coriolis parameter f at a reference latitude ($f = f_0 + \beta y$), N_B is the Brunt–Väisälä frequency (**Section 10.13.2.4**), and ρ_0 is a height-dependent zonal and time-mean density, that is, $\rho_s \exp(-z/H)$, where H is density scale height and ρ_s a reference density. Just as barotropic instability was caused by a change of sign of $\partial\eta/\partial y$ in eqn [116], so instability arises when $\partial \overline{q}_s/\partial y$ changes sign. When the third term $(1/\rho_0)\partial/\partial z(\rho_0 B \partial \overline{u}/\partial z)$ dominates, a different kind of instability emerges. This is baroclinic instability, which is when flow is driven by the extraction of potential energy from a latitudinal temperature gradient, causing large-scale eddies. If we use the barometric law $\rho_0 = \rho_s \exp(-z/H)$ and differentiate the last term of eqn [120] w.r.t. z using the product rule, we see that this term depends on the vertical wind shear:

$$\left(\frac{1}{\rho_0}\right)\frac{\partial}{\partial z}\left[(\rho_0 B)\frac{\partial \overline{u}}{\partial z}\right] \approx \left(\frac{B}{\rho_0}\right)\left(-\left(\frac{\rho_0}{H}\right)\frac{\partial \overline{u}}{\partial z} + \rho_0 \cancel{\frac{\partial^2 \overline{u}}{\partial z^2}}\right) \propto -\left(\frac{1}{H}\right)\frac{\partial \overline{u}}{\partial z} \qquad [121]$$

Thus, baroclinic instability is most commonly found near the regions of strong vertical shear of the mean zonal wind. On Earth, it occurs when $\partial \overline{u}/\partial z$ exceeds $\sim 1\ \text{m s}^{-1}\ \text{km}^{-1}$. This threshold is small and less than the wind shear predicted from the thermal wind equation with a typical horizontal gradient induced by the spread of sunlight over midlatitudes. Consequently, baroclinic instability is generally predicted on rapidly rotating rocky planets at the appropriate latitudes and occurs on Mars, making associated baroclinic waves (Banfield et al., 2004; Collins and James, 1995; Hinson et al., 2012; Read and Lewis, 2003; Wilson et al., 2002). The net thermal effect of 3-D baroclinic eddies is to transport warm low-latitude air poleward and up and cold high-latitude air equatorward and down. For this reason, the process has been called stantwise (or sloping) convection (Hide and Mason, 1975). As described in **Section 10.13.5.4**, eddies can drive jet streams. However, baroclinic instability should be less important on slowly rotating planets with small horizontal temperature gradients.

The horizontal length scale of the most unstable eddies produced by baroclinic instability is the Rossby radius of deformation:

$$r_{Ro} = \frac{N_B H}{f_0} \qquad [122]$$

where H is the scale height, N_B the Brunt–Vaisala frequency, and f_0 a reference Coriolis parameter. The Rossby radius can be understood as follows. Imagine air rising vertically through a height H and then spreading out perhaps when it reaches the tropopause. Initially, the spreading gas will not 'feel' the effects of the planetary rotation, but after a certain distance, its lateral spread will be converted to rotational motion around a core. This distance is roughly the Rossby radius. In the Earth's midlatitude troposphere, $r_{Ro} \sim 700$ km, using $f_0 \sim 10^{-4}\ \text{s}^{-1}$, $H \sim 7$ km, and $N_B \sim 10^{-2}\ \text{s}^{-1}$. Perhaps as a consequence, only one eddy-driven jet can fit in the Earth's midlatitudes (Schneider and Walker, 2006). On Jupiter, $r_{Ro} \sim [(0.003\text{–}0.02) \times 27]/[2 \times (2\pi/35730) \times \sin(45°)] \approx 325\text{–}2200$ km using a range of N_B values (Magalhaes et al., 2002; Shetty and Marcus, 2010), which is small compared to Jupiter's vast size.

10.13.5.6 Deep and Shallow Atmospheres on Giant Planets

The pattern of strong eastward jets at the equator on Jupiter and Saturn, with alternatively eastward and westward jets at higher latitudes is a puzzle of atmospheric dynamics. Historically, there have been two approaches to explain the jet structure (Vasavada and Showman, 2005):

(i) The observed wind systems are occurring in a shallow layer overlying a horizontally uniform but convective deep interior (Williams, 1978).
(ii) The observed winds somehow reflect the dynamic activity of an active deep atmosphere, and the jets probably represent very deeply rooted systems (Busse, 1976, 1983).

In the first scenario, shallow-layer models can produce eddies (Cho and Polvani, 1996; Huang and Robinson, 1998) – sometimes associated with the moist convection (Lian and Showman, 2010) – that force alternating zonal jets, though many models produce westward jets at the equator. In the second scenario, circulations extend deeply into the interior. The deep interior opacities of giant planets should lead to nearly adiabatic and convective conditions (Nutzman et al., 2009). The jets themselves could perhaps result from the surface manifestation of large circulating cells in systematic alignment with the jets. Galileo probe measurements of a vertical wind profile on Jupiter show nearly uniform strong zonal

winds to great depth from 5 to 20 bar, but consistent with temperature measurements interpreted in the light of the thermal wind equation, these measurements show decay of wind strength upward in the uppermost troposphere (Atkinson et al., 1997). This wind profile supports the deep circulation camp, but because the profile may not be representative, it does not prove it. The probe descended in a 'hot spot' of high IR emission at 5 μm, which emerges from deeper, unusually cloud-free parts of the troposphere (Ortiz et al., 1998).

On Uranus and Neptune (the ice giants), a broad band of strong retrograde winds exist at low latitudes with a single high-latitude prograde jet in each hemisphere (**Figure 14**). Moreover, Neptune's retrograde jet speed exceeds 500 m s^{-1}, the fastest in the solar system. It is unclear why there are such significant circulation differences between the gas giants and Uranus and Neptune. Possibly, convection is weaker on the ice giants and unable to drive the eddies needed for the kind of jet structure seen on the giant planets or only operates to drive prograde jets at high latitudes. Perhaps, a surprising result is that the jet speeds increase going outward in the solar system, which is counterintuitive if one assumes that incident sunlight is important. This puzzle has been called Suomi's paradox after Vern Suomi, who first pointed it out and is not resolved.

10.13.5.7 Buoyancy Waves and Thermal Tides

Besides planetary waves, another class of atmospheric waves that are very important for circulation are propagating buoyancy waves (also called internal gravity waves), which correspond to oscillations primarily in the vertical plane when the square of the Brunt–Väisälä frequency exceeds zero, $N_B^2 > 0$. Given that N_B^2 has the same sign as static stability (eqn [17]), this situation is common in atmospheres. As discussed in **Section 10.13.2.4**, the restoring force of the mean vertical buoyancy gradient acts on vertically displaced fluid parcels to cause oscillations. It is possible to derive a wave equation for buoyancy waves called the Taylor–Goldstein equation and find solutions (e.g., Holton, 2004, pp. 196–201; Nappo, 2002). In the x-z plane, wave perturbations in density, vertical velocity, pressure, and potential temperature vary as $\sim e^{(z/2H)} e^{i(kx-\omega t)}$, where the factor $\exp(z/2H)$ accounts for the vertical variation of atmospheric density, where H is the density scale height.

Such waves are generated mainly by flow over topography and large-scale flow associated with convective cells. Wave amplitude varies with density as $\rho^{-1/2}$, so amplitudes grow upward as density decreases, causing some waves to break aloft, in crude analogy to ocean waves breaking on a beach. Theory shows that buoyancy waves can also be absorbed at lower levels, the so-called critical levels, if the phase speed of the wave c equals the mean zonal wind, that is, $c = \overline{u}$ (Forbes, 2002; Nappo, 2002). Otherwise, the waves continue to propagate vertically and break well above these critical levels. Thus, for example, if an eastward jet exists at altitude, buoyancy waves propagating upward with eastward phase speeds will be absorbed and are said to be filtered out, whereas waves with westward or near-zero relative speed (the speed of the underlying surface) will penetrate higher, break, and dump their westward momentum (Forbes, 2002). The result is that these breaking waves diminish eastward zonal winds aloft, which is called closing off the winds – that is, a jet exists below the breaking altitude and winds diminish upward from the core of a jet. This process is important for shaping jets in general, acts as a drag on the zonal mean flow, and makes upper atmospheres 'feel' the rotation of the underlying planet. Thermal wind balance (eqn [104]) applied to the effect of waves closing off the jets and declining wind speed with altitude is associated with a reverse equator-to-pole temperature gradient in the Earth's stratosphere. In turn, through thermodynamic balance (eqn [107]), this is connected with ascent and descent, where tropospheric air rises into the tropical stratosphere and subsides at the poles – the Brewer–Dobson circulation (Brewer, 1949; Dobson, 1956) – which carries stratospheric ozone from equator to poles.

An important and special class of large-scale buoyancy waves are atmospheric thermal tides, which are global oscillations in pressure and wind fields caused by the heating and cooling associated with the apparent motion of the Sun over a planet (Chapman and Lindzen, 1970). The wave periods are fractions of a solar day, so that the wave with a solar day length is a diurnal tide, while that with half a day length is a semidiurnal tide. Furthermore, the tides divide into migrating tides that move around the planet with the Sun's apparent motion and nonmigrating tides that do not, which are modes generated by migrating tides interacting with planet-scale topography.

A key theoretical result is that global-scale thermal tidal waves only propagate vertically if their angular frequency ω exceeds the Coriolis parameter, that is, $\omega > f$ (Lindzen, 1971). A diurnal tide with $\omega = \Omega$ cannot propagate vertically beyond 30° latitude because $f = 2\Omega\sin(30°) = \Omega$ at that latitude. Thus, on Earth and Mars, the diurnal tide acts as a buoyancy wave only between latitudes ±30° where it propagates vertically and its amplitude increases with height, whereas the semidiurnal tide is a global buoyancy wave from pole to pole with very large vertical wavelength ~100 km. Migrating and nonmigrating tides have been detected on Earth and Mars, and migrating tides have been observed on Venus (Migliorini et al., 2012; Pechmann and Ingersoll, 1984; Peralta et al., 2012; Schofield and Taylor, 1983). In the case of Venus, which rotates every 243 Earth days, the thermal tides in the middle atmosphere reflect atmospheric rotation, clearly not the rotation of the solid planet. The atmosphere near the cloud tops at ~65–70 km altitude rotates at ~110 m s^{-1} past the subsolar point, establishing waves that migrate with the Sun but move opposite to the direction of the mean flow (Gierasch et al., 1997).

Tides are important in atmospheres. On Mars, feedback between tidal surface winds and lifting of dust high in the atmosphere may contribute to the generation of planet encircling dust storms (Leovy et al., 1973), while breaking of tides may be important for mixing atmospheric constituents. On Venus, tides may be implicated in superrotation, which we discuss next.

10.13.5.8 Superrotation

Superrotation is where the equatorial zonal wind speed is persistently greater than the speed of the underlying rotating planet, which applies to Venus (Belton et al., 1991; Peralta et al., 2007), Titan (Bird et al., 2005; Kostiuk et al., 2006), Jupiter (Ingersoll et al., 2004), and Saturn (Del Genio et al., 2009). The angular momentum of a planet's surface is maximal

at the equator, and so, the superrotation cannot be maintained by some simple transport of angular momentum (Hide, 1969, 1970), as, for example, occurs from equator to high latitudes in Hadley cells. Also, since viscosity should tend to dissipate high equatorial winds, superrotation is puzzling.

Generally, it has been proposed that equatorial superrotation might be maintained by eddies with the kind of 'upgradient' transport discussed in **Section 10.13.5.4** (e.g., Gierasch, 1975) (although gravitational torque acting on the atmosphere is an alternative idea (Gold and Soter, 1971; Schubert, 1983)). Eddies could carry prograde momentum systematically toward the jet region, either by motions in the vertical plane with upward and zonal velocities correlated below the maximum wind level and downward and zonal velocities correlated above the maximum or by motions in the horizontal plane with equatorward and prograde winds correlated (Schneider and Liu, 2009). If the former case applies, buoyancy waves may be implicated; in the latter case, Rossby waves. Either way, generation of wave activity by some process near the equatorial wind maximum would be required.

Attempts to model superrotation on Venus have been made with varying degrees of success, and models that work have generally assumed background superrotation below the cloud base (e.g., Del Genio and Zhou, 1996; Hollingsworth et al., 2007; Lebonnois et al., 2010; Yamamoto and Takahashi, 2009). The waves near the superrotating level on Venus may be atmospheric tides (Newman and Leovy, 1992). It is unlikely to be a coincidence that the peak winds occur near a maximum in the vertical profile of solar heating maximum (**Figure 17**) (Gierasch et al., 1997; Leovy, 1987; Pechmann and Ingersoll, 1984; Schofield and Taylor, 1983). However, the underlying physics of equatorial superrotation remains an area of research (Mitchell and Vallis, 2010).

10.13.6 Outstanding Questions

Although the Space Age has greatly improved our knowledge of the atmospheres of the solar system, including the Earth, a long list of unsolved problems remains. Planetary atmospheres as a whole are very undersampled, with no in situ measurements at all for Saturn, Uranus, and Neptune, a possibly unrepresentative descent profile from Jupiter, poor knowledge of the atmosphere of Venus below the clouds, and barely any measurements of winds in all extraterrestrial atmospheres. Even on Earth, matters of basic enquiry remain. These include, for example, how the Hadley circulation interacts with eddies (Levine and Schneider, 2011), how stratospheric circulation and

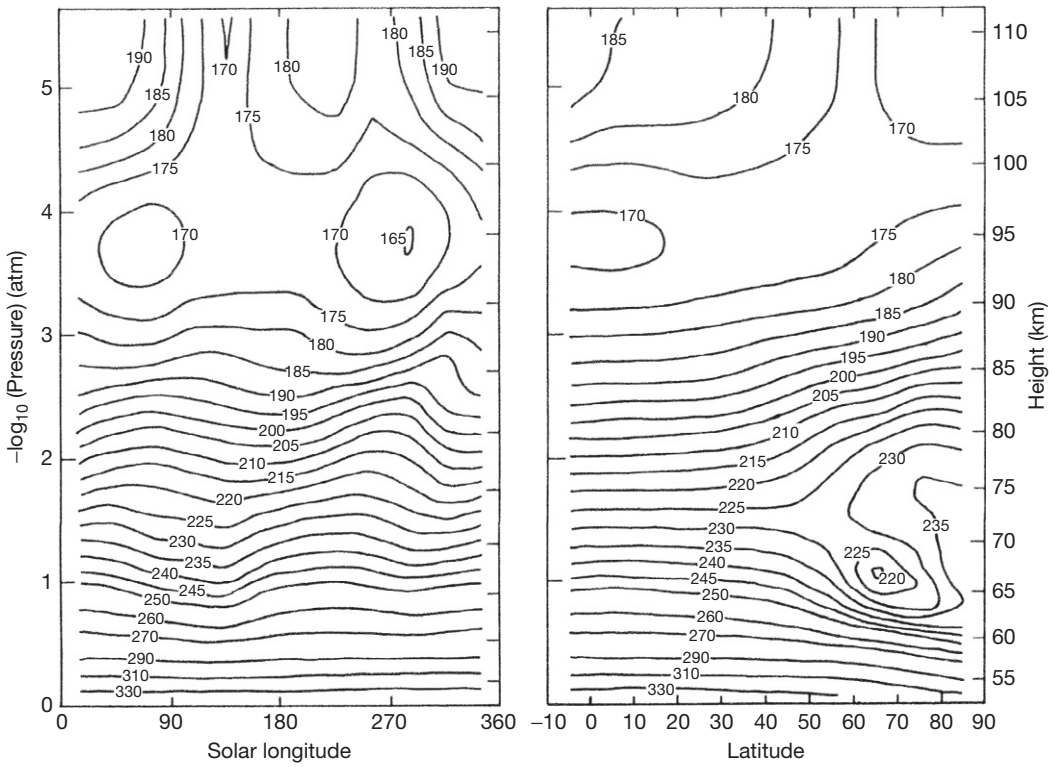


Figure 17 Cross sections of temperature (K) in the atmosphere of Venus above 55 km altitude (Schofield and Taylor, 1983). On the left, the temperature field is averaged between 0 and30°N and shown in Sun-following coordinates where zero longitude is noon and 180° is the antisolar point. There is a minimum temperature of 170 K at 95 km altitude, above which the equator-to-pole temperature gradient is reversed. On the right, temperatures are shown as a zonal mean cross section. From above the cloud tops (70 km) to 95 km, a semidiurnal variation is revealed by waves in the temperature contours with a phase that moves in an eastward direction at higher altitude (left). The phase in the solar longitude plot on the left suggests a Sun-following eastward-propagating oscillation forced at lower altitudes. Gierasch PJ, Goody RM, Young RE, et al. (1997) The general circulation of the Venus atmosphere: An assessment. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 459–500. Tucson, AZ: University of Arizona Press.

chemistry affects the troposphere (Thompson et al., 2011), and the response of cloud formation (Boucher et al., 2013) and the stability of ice sheets to climate forcing (McNeall et al., 2011).

After the Earth, Mars has the most observed atmosphere. On Mars, global dust storms occur in some Martian years and not others (Martin and Zurek, 1993). But we do not know why. What are the mechanisms for storm initiation, growth, and decay and how are these storms connected to circulation features such as tides and planetary waves? Also, the role of clouds and surface exchange in the water cycle on Mars and its interaction with the dust cycle remains to be elucidated.

On Venus, winds and temperature fields in the lower atmosphere are essential for understanding the circulation and yet are poorly measured or understood. Clouds play such an important role for Venus' climate that further elucidation of their size distribution, shapes, and composition and the distribution of minor sulfur and chlorine species would also help define chemistry, climate, and dynamics. Also, many reactions relevant to chemistry on Venus that involve elemental sulfur lack laboratory data.

We do not have a fundamental understanding of the differences between the giant-planet atmospheric circulations. There are two main puzzles. The first is the different jet structures on Uranus and Neptune compared with those on Jupiter and Saturn. The latter have several jets in each hemisphere (Jupiter has 5–6 eastward jets and Saturn has 4), whereas the ice giants have only one jet (**Figure 14**). Whereas the equatorial jets are retrograde on the ice giants, they are prograde on the gas giants. Also, the fastest jets are on Neptune, and although Uranus is tipped on its side, its meridional profile of zonal winds is much like Neptune's and vaguely like the Earth's. Clearly, incident sunlight plays less of an overall role than rotation, while obliquity seems strangely unimportant for the jet structure. The second main puzzle is equatorial superrotation on Jupiter and Saturn. The superrotation is so strong on these gas giants that it is probably very deep. Vertical wind shear is required for the buoyancy wave mechanism that may be important for Venus yet appears lacking based on Galileo probe results (**Section 10.13.5.8**). So unlike on Venus, there is no clear mechanism for generating sufficiently strong buoyancy wave momentum fluxes. The horizontal shear zones associated with superrotation on Jupiter and Saturn are much wider than the scales of eddies that might transport momentum horizontally to drive superrotation. What drives superrotation on Saturn and Jupiter? And why is it absent on Uranus and Neptune, which have strong equatorial easterlies?

Finally, this chapter omitted atmospheric evolution for reasons of space. But it is clear from the history of the research literature in planetary science that initial assumptions of Earth-like conditions on the other rocky bodies, Venus and Mars, proved to be drastically erroneous. The reasons are related to great differences in how these bodies evolved, which is another story (Catling and Kasting, in press).

Acknowledgments

Support for this work included NASA Mars Fundamental Research Program Grant NNX10AN67G and Exobiology/Astrobiology Grant NNX10AQ90G awarded to the author and NASA Astrobiology Institute's Virtual Planetary Laboratory award under solicitation no. NNH05ZDA001C. The author also thanks the late Conway Leovy for helping improve his knowledge of planetary atmospheres and Tyler Robinson for a helpful review.

References

Abe Y (2011) Protoatmospheres and surface environment of protoplanets. *Earth, Moon, and Planets* 108: 9–14.

Andrews DG (2010) *An Introduction to Atmospheric Physics*, 2nd edn. New York: Cambridge University Press.

Appleby JF (1986) Radiative convective equilibrium models of Uranus and Neptune. *Icarus* 65: 383–405.

Appleby JF and Hogan JS (1984) Radiative-convective equilibrium models of Jupiter and Saturn. *Icarus* 59: 336–366.

Armstrong BH (1968) Theory of the diffusivity factor for atmospheric radiation. *Journal of Quantitative Spectroscopy and Radiative Transfer* 8: 1577–1599.

Atkins PW and De Paula J (2010) *Physical Chemistry*, 9th edn. New York: W.H. Freeman.

Atkinson DH, Ingersoll AP, and Seiff A (1997) Deep winds on Jupiter as measured by the Galileo probe. *Nature* 388: 649–650.

Atreya SK, Lorenz RD, and Waite JH (2009) Volatile origin and cycles: Nitrogen and methane. In: Brown RH, Lebreton J-P, and Waite JH (eds.) *Titan from Cassini-Huygens*, pp. 77–99. New York: Springer.

Atreya SK, Mahaffy PR, Niemann HB, Wong MH, and Owen TC (2003) Composition and origin of the atmosphere of Jupiter – An update, and implications for the extrasolar giant planets. *Planetary and Space Science* 51: 105–112.

Baker NL and Leovy CB (1987) Zonal winds near Venus cloud top level: A model study of the interaction between the zonal mean circulation and the semidiurnal tide. *Icarus* 69: 202–220.

Banfield D, Conrath BJ, Gierasch PJ, Wilson RJ, and Smith MD (2004) Traveling waves in the Martian atmosphere from MGS TES Nadir data. *Icarus* 170: 365–403.

Belton MJS, Gierasch PJ, Smith MD, et al. (1991) Images from Galileo of the Venus cloud deck. *Science* 253: 1531–1536.

Bent HA (1965) *The Second Law: An Introduction to Classical and Statistical Thermodynamics*. New York: Oxford University Press.

Berner RA (2004) *The Phanerozoic Carbon Cycle*. New York: Oxford University Press.

Bird MK, Allison M, Asmar SW, et al. (2005) The vertical profile of winds on Titan. *Nature* 438: 800–802.

Boucher O, Randall D, Artaxo P, et al. (2013) Clouds and aerosols. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, and Midgley PM (eds.) *Climate Change 2013: The Physical Science Basis*, pp. 571–657. New York: Cambridge University Press.

Brewer AW (1949) Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quarterly Journal of the Royal Meteorological Society* 75: 351–363.

Brown GN and Ziegler WT (1979) Vapor pressure and heats of sublimation of liquids and solids of interest in cryogenics below 1-atm pressure. In: Timmerhaus K and Snyder HA (eds.) *Advances in Cryogenic Engineering*, pp. 662–670. New York: Plenum Press.

Buick R (2007) The earliest records of life on Earth. In: Sullivan WT and Baross J (eds.) *Planets and Life: The Emerging Science of Astrobiology*, pp. 237–264. Cambridge: Cambridge University Press.

Bullock MA and Grinspoon DH (2001) The recent evolution of climate on Venus. *Icarus* 150: 19–37.

Busse FH (1976) Simple model of convection in Jovian atmosphere. *Icarus* 29: 255–260.

Busse FH (1983) A model of mean zonal flows in the major planets. *Geophysical and Astrophysical Fluid Dynamics* 23: 153–174.

Carlson RW (1999) A tenuous carbon dioxide atmosphere on Jupiter's moon Callisto. *Science* 283: 820–821.

Catling DC (2014) The great oxidation event transition. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, pp. 177–195. New York: Elsevier.

Catling DC, Claire MW, Zahnle KJ, et al. (2010) Atmospheric origins of perchlorate on Mars and in the Atacama. *Journal of Geophysical Research* 115: E00E11.

Catling DC and Kasting JF (in press) *Atmospheric Evolution on Inhabited and Lifeless Worlds*. New York: Cambridge University Press.

Catling DC and Zahnle KL (2009) The planetary air leak. *Scientific American* 300: 36–43.

Chamberlain JW and Hunten DM (1987) *Theory of Planetary Atmospheres*. Orlando, FL: Academic Press.

Chambers JE (2014) Planet formation. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, pp. 55–72. New York: Elsevier.

Chapman GA, Cookson AM, and Preminger DG (2012) Comparison of TSI from SORCE TIM with SFO ground-based photometry. *Solar Physics* 276: 35–41.

Chapman S and Lindzen RS (1970) *Atmospheric Tides: Thermal and Gravitational*. New York, NY: Gordon and Breach.

Cho JYK and Polvani LM (1996) The morphogenesis of bands and zonal winds in the atmospheres on the giant outer planets. *Science* 273: 335–337.

Colburn D, Pollack J, and Haberle R (1989) Diurnal variations in optical depth at Mars. *Icarus* 79: 159–189.

Collins M and James IN (1995) Regular baroclinic transient waves in a simplified global circulation model of the martian atmosphere. *Journal of Geophysical Research* 100: 14421–14432.

Cook KH (2004) Hadley circulation dynamics. In: Diaz HF and Bradley RS (eds.) *The Hadley Circulation: Past, Present and Future*, pp. 61–83. Dordrecht: Kluwer.

Courtin R, Mckay CP, and Pollack J (1992) The greenhouse effect within the Solar System. *Recherche* 23: 542–549.

Crisp D (1986) Radiative forcing of the Venus mesosphere. I – Solar fluxes and heating rates. *Icarus* 67: 484–514.

Crisp D (1989) Radiative forcing of the Venus mesosphere. II – Thermal fluxes, cooling rates, and radiative equilibrium temperatures. *Icarus* 77: 391–413.

Crisp D and Titov D (1997) The thermal balance of the Venus atmosphere. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 353–384. Tucson, AZ: University of Arizona Press.

Crutzen PJ (1976) Possible importance of CSO for sulfate layer of stratosphere. *Geophysical Research Letters* 3: 73–76.

Crutzen PJ (1979) Role of NO and NO_2 in the chemistry of the troposphere and stratosphere. *Annual Review of Earth and Planetary Sciences* 7: 443–472.

Curtis AR and Goody RM (1956) Thermal radiation in the upper atmosphere. *Proceedings of the Royal Society of London A* 236: 193–206.

Del Genio AD, Achterberg RK, Baines KH, et al. (2009) Saturn atmospheric structure and dynamics. In: Dougherty M, Esposito L, and Krimigis S (eds.) *Saturn from Cassini-Huygens*, pp. 113–160. New York: Springer.

Del Genio AD and Suozzo RJ (1987) A comparative study of rapidly and slowly rotating dynamic regimes in a terrestrial general circulation model. *Journal of the Atmospheric Sciences* 44: 973–986.

Del Genio AD and Zhou W (1996) Simulations of superrotation on slowly rotating planets: Sensitivity to rotation and initial condition. *Icarus* 120: 332–343.

Dima IM and Wallace JM (2003) On the seasonality of the Hadley cell. *Journal of the Atmospheric Sciences* 60: 1522–1527.

Dobson GMB (1956) Origin and distribution of the polyatomic molecules in the atmosphere. *Proceedings of the Royal Society of London A* 236: 187–193.

Fegley B (2014) Venus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, pp. 127–148. New York: Elsevier.

Forbes JM (2002) Wave coupling in terrestrial planetary atmospheres. In: Mendillo M, Nagy AF, and Waite JH (eds.) *Atmospheres in the Solar System: Comparative Aeronomy*, pp. 171–190. Washington, D.C.: American Geophysical Union.

Gierasch PJ (1975) Meridional circulation and maintenance of Venus atmospheric rotation. *Journal of the Atmospheric Sciences* 32: 1038–1044.

Gierasch PJ, Goody RM, Young RE, et al. (1997) The general circulation of the Venus atmosphere: An assessment. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 459–500. Tucson, AZ: University of Arizona Press.

Gold T and Soter S (1971) Atmospheric tides and 4-day circulation on Venus. *Icarus* 14: 16–20.

Goldblatt C, Robinson TD, Zahnle KJ, and Crisp D (2013) Low simulated radiation limit for runaway greenhouse climates. *Nature Geoscience* 6: 661–667.

Goody RM and Yung YL (1989) *Atmospheric Radiation: Theoretical Basis*, 2nd edn. New York: Oxford University Press.

Greeley R, Bender KC, Saunders RS, Schubert G, and Weitz C (1997) Aeolian processes and features on Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 547–589. Tucson, AZ: University of Arizona Press.

Griffith CA and Zahnle K (1995) Influx of cometary volatiles to planetary Moons: The atmospheres of 1000 possible Titans. *Journal of Geophysical Research* 100: 16907–16922.

Guillot T (1995) Condensation of methane, ammonia, and water and the inhibition of convection in giant planets. *Science* 269: 1697–1699.

Haberle RM, McKay CP, Gwynne OE, et al. (1993a) Atmospheric effects on the utility of solar power on Mars. In: Lewis J and Matthews MS (eds.) *Resources of Near-Earth Space*, pp. 845–885. Tucson, AZ: University of Arizona Press.

Haberle RM, Pollack JB, Barnes JR, et al. (1993b) Mars atmospheric dynamics as simulated by the NASA Ames General Circulation Model. 1. The zonal-mean circulation. *Journal of Geophysical Research* 98: 3093–3123.

Hadley G (1735) Concerning the cause of the general trade-winds. *Philosophical Transactions of the Royal Society of London* 39: 58–62.

Hall DT, Feldman PD, McGrath MA, and Strobel DF (1998) The far-ultraviolet oxygen airglow of Europa and Ganymede. *Astrophysical Journal* 499: 475–481.

Han TT, Ping JS, and Zhang SJ (2011) Global features and trends of the tropopause derived from GPS/CHAMP RO data. *Science China – Physics Mechanics & Astronomy* 54: 365–374.

Hanel RA, Conrath BJ, Herath LW, Kunde VG, and Pirraglia JA (1981) Albedo, internal heat, and energy balance of Jupiter: Preliminary results of the Voyager infrared investigation. *Journal of Geophysical Research* 86: 8705–8712.

Hanel RA, Conrath BJ, Kunde VG, et al. (1972) The Nimbus 4 infrared spectroscopy. Experiment, I: Calibrated thermal emission spectra. *Journal of Geophysical Research* 77: 2629–2641.

Hansen CJ, Shemansky DE, Esposito LW, et al. (2011) The composition and structure of the Enceladus plume. *Geophysical Research Letters* 38.

Hartmann DL, Klein Tank AMG, Rusticucci M, et al. (2013) Observations: Atmosphere and surface. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, and Midgley PM (eds.) *Climate Change 2013: The Physical Science Basis*, pp. 159–254. New York: Cambridge University Press.

Held IM (1975) Momentum transport by quasi-geostrophic eddies. *Journal of the Atmospheric Sciences* 32: 1494–1497.

Held IM (2000) General circulation of the atmosphere. In: Thiffeault J-L (ed.) *2000 Program in Geophysical Fluid Dynamics*, pp. 1–54. Woods Hole, MA: Woods Hole Oceanographic Institution (http://www.whoi.edu/page.do?pid=13076).

Held IM and Hou AY (1980) Non-linear axially-symmetric circulations in a nearly inviscid atmosphere. *Journal of the Atmospheric Sciences* 37: 515–533.

Heng K, Hayek W, Pont F, and Sing DK (2012) On the effects of clouds and hazes in the atmospheres of hot Jupiters: Semi-analytical temperature-pressure profiles. *Monthly Notices of the Royal Astronomical Society* 420: 20–36.

Hide R (1969) Dynamics of atmospheres of major planets with an appendix on viscous boundary layer at rigid bounding surface of an electrically-conducting rotating fluid in presence of a magnetic field. *Journal of the Atmospheric Sciences* 26: 841–853.

Hide R (1970) Equatorial jets in planetary atmospheres. *Nature* 225: 254–255.

Hide R and Mason PJ (1975) Sloping convection in a rotating fluid. *Advances in Physics* 24: 47–100.

Hinson DP, Wang HQ, and Smith MD (2012) A multi-year survey of dynamics near the surface in the northern hemisphere of Mars: Short-period baroclinic waves and dust storms. *Icarus* 219: 307–320.

Hoinka KP (1998) Statistics of the global tropopause pressure. *Monthly Weather Review* 126: 3303–3325.

Hollingsworth JL and Barnes JR (1996) Forced stationary planetary waves in Mars's winter atmosphere. *Journal of the Atmospheric Sciences* 53: 428–448.

Hollingsworth JL, Young RE, Schubert G, Covey C, and Grossman AS (2007) A simple-physics global circulation model for Venus: Sensitivity assessments of atmospheric superrotation. *Geophysical Research Letters* 34.

Holton JR (2004) *An Introduction to Dynamic Meteorology, International geophysics series*, 4th edn. New York: Elsevier Academic Press.

Houben H (2013) Thermophysical fluid dynamics: The key to the structures of fluid objects. *AGU Fall Meeting Conference, San Francisco December 9–13 2013*, abstract #P21C-1743.

Houghton JT (2002) *The Physics of Atmospheres*, 3rd edn. New York: Cambridge University Press.

Hu RY, Seager S, and Bains W (2012) Photochemistry in terrestrial exoplanet atmospheres. I. Photochemistry model and benchmark cases. *Astrophysical Journal* 761.

Hu RY, Seager S, and Bains W (2013) Photochemistry in terrestrial exoplanet atmospheres. II. H_2S and SO_2 photochemistry in anoxic atmospheres. *Astrophysical Journal* 769.

Huang HP and Robinson WA (1998) Two-dimensional turbulence and persistent zonal jets in a global barotropic model. *Journal of the Atmospheric Sciences* 55: 611–632.

Hunten DM (1973) The escape of light gases from planetary atmospheres. *Journal of the Atmospheric Sciences* 30: 1481–1494.

Ingersoll AP (1969) The runaway greenhouse: A history of water on Venus. *Journal of the Atmospheric Sciences* 26: 1191–1198.

Ingersoll AP (1990) Atmospheric dynamics of the outer planets. *Science* 248: 308–315.

Ingersoll AP (2013) *Planetary Climates, Princeton Primers in Climate*. Princeton, NJ: Princeton University Press.

Ingersoll AP, Dowling TE, Gierasch PJ, et al. (2004) Dynamics of Jupiter's atmosphere. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites, and Magnetosphere*. New York: Cambridge University Press.

Irwin P (2009) *Giant Planets of our Solar System: Atmospheres, Composition, and Structure*, 2nd edn. Chichester, UK: Springer.

Jaumann R, Kirk RL, Lorenz RD, et al. (2009) Geology and surface processes on Titan. In: Brown RH, Lebreton J-P, and Waite JH (eds.) *Titan from Cassini-Huygens*, pp. 75–140. New York: Springer.

Johnson RE, Burger MH, Cassidy TA, Leblanc F, Marconi M, and Smyth WH (2009) Composition and detection of Europa's sputter-induced atmosphere. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 507–527. Tucson, AZ: University of Arizona Press.

Johnson RE, Volkov AN, and Erwin JT (2013) Molecular-kinetic simulations of escape from the ex-planet and exoplanets: Criterion for transonic flow. *Astrophysical Journal Letters* 768.

Joshi MM, Lewis SR, Read PL, and Catling DC (1994) Western boundary currents in the atmosphere of Mars. *Nature* 367: 548–551.

Joshi MM, Lewis SR, Read PL, and Catling DC (1995) Western boundary currents in the Martian atmosphere: Numerical simulations and observational evidence. *Journal of Geophysical Research* 100: 5485–5500.

Kahn RA, Martin TZ, and Zurek RW (1992) The martian dust cycle. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1017–1053. Tucson, AZ: University of Arizona Press.

Karl TR and Trenberth KE (2003) Modern global climate change. *Science* 302: 1719–1723.

Kasting JF (1988) Runaway and moist greenhouse atmospheres and the evolution of Earth and Venus. *Icarus* 74: 472–494.

Kasting JF (1991) CO_2 condensation and the climate of early Mars. *Icarus* 94: 1–13.

Kasting JF and Ackerman TP (1986) Climatic consequences of very high CO_2 levels in the earth's early atmosphere. *Science* 234: 1383–1385.

Kirschke S, Bousquet P, Ciais P, et al. (2013) Three decades of global methane sources and sinks. *Nature Geoscience* 6: 813–823.

Kliore AJ, Anabtawi A, Herrera RG, et al. (2002) Ionosphere of Callisto from Galileo radio occultation observations. *Journal of Geophysical Research* 107.

Kopp G and Lean JL (2011) A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters* 38.

Kostiuk T, Livengood TA, Sonnabend G, et al. (2006) Stratospheric global winds on Titan at the time of Huygens descent. *Journal of Geophysical Research* 111: E07S03.

Krasnopolsky VA (2010) The photochemical model of Titan's atmosphere and ionosphere: A version without hydrodynamic escape. *Planetary and Space Science* 58: 1507–1515.

Krasnopolsky VA (2011) Atmospheric chemistry on Venus, Earth, and Mars: Main features and comparison. *Planetary and Space Science* 59: 952–964.

Krasnopolsky VA and Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294: 1914–1917.

Krasnopolsky VA and Lefevre F (2013) Chemistry of the atmospheres of Mars, Venus, and Titan. In: Mackwell SJ, Simon-Miller AA, Harder JW, and Bullock MA (eds.) *Comparative Climatology of Terrestrial Planets*, pp. 231–275. Tucson, AZ: University of Arizona Press.

Krasnopolsky VA and Pollack JB (1994) H_2O-H_2SO_4 system in Venus' clouds and OCS, CO, and H_2SO_4 profiles in Venus' troposphere. *Icarus* 109: 58–78.

Lacis AA, Hansen JE, Russell GL, Oinas V, and Jonas J (2013) The role of long-lived greenhouse gases as principal LW control knob that governs the global surface temperature for past and future climate change. *Tellus Series B-Chemical and Physical Meteorology* 65: 19734.http://dx.doi.org/10.3402/tellusb.v65i0.19734.

Lacis AA, Schmidt GA, Rind D, and Ruedy RA (2010) Atmospheric CO_2: Principal control knob governing Earth's temperature. *Science* 330: 356–359.

Lammer H (2013) *Origin and Evolution of Planetary Atmospheres: Implications for Habitability*. New York: Springer.

Lammer H, Kasting JF, Chassefiere E, Johnson RE, Kulikov YN, and Tian F (2008) Atmospheric escape and evolution of terrestrial planets and satellites. *Space Science Reviews* 139: 399–436.

Lavvas PP, Coustenis A, and Vardavas IM (2008a) Coupling photochemistry with haze formation in Titan's atmosphere, part I: Model description. *Planetary and Space Science* 56: 27–66.

Lavvas PP, Coustenis A, and Vardavas IM (2008b) Coupling photochemistry with haze formation in Titan's atmosphere, Part II: Results and validation with Cassini/Huygens data. *Planetary and Space Science* 56: 67–99.

Lebonnois S, Hourdin F, Eymet V, Crespin A, Fournier R, and Forget F (2010) Superrotation of Venus' atmosphere analyzed with a full general circulation model. *Journal of Geophysical Research* 115.

Lee S and Kim HK (2003) The dynamical relationship between subtropical and eddy-driven jets. *Journal of the Atmospheric Sciences* 60: 1490–1503.

Lefevre F and Forget F (2009) Observed variations of methane on Mars unexplained by known atmospheric chemistry and physics. *Nature* 460: 720–723.

Leighton RB and Murray BC (1966) Behaviour of carbon dioxide and other volatiles on Mars. *Science* 153: 136–144.

Lellouch E, de Bergh C, Sicardy B, Kaufl HU, and Smette A (2011) The tenuous atmospheres of Pluto and Triton explored by CRIRES on the VLT. *ESO Messenger* 145: 20–23.

Lellouch E, Paubert G, Moses JI, Schneider NM, and Strobel DF (2003) Volcanically emitted sodium chloride as a source for Io's neutral clouds and plasma torus. *Nature* 421: 45–47.

Leovy CB (1987) Zonal winds near Venus cloud top level: An analytic model of the equatorial wind speed. *Icarus* 69: 193–201.

Leovy CB, Zurek RW, and Pollack JB (1973) Mechanisms for Mars dust storms. *Journal of the Atmospheric Sciences* 30: 749–762.

Levine XJ and Schneider T (2011) Response of the Hadley circulation to climate change in an aquaplanet GCM coupled to a simple representation of ocean heat transport. *Journal of the Atmospheric Sciences* 67: 769–783.

Levy H (1971) Normal atmosphere – Large radical and formaldehyde concentrations predicted. *Science* 173: 141–143.

Lian Y and Showman AP (2010) Generation of equatorial jets by large-scale latent heating on the giant planets. *Icarus* 207: 373–393.

Liang MC, Lane BF, Pappalardo RT, Allen M, and Yung YL (2005) Atmosphere of Callisto. *Journal of Geophysical Research – Planets* 110.

Lindemann TA and Dobson GMB (1923) A theory of meteors, and the density and temperature of the outer atmosphere to which it leads. *Proceedings of the Royal Society of London A* 102: 411–437.

Lindzen RS (1971) Atmospheric tides. In: Reid WH (ed.) *Mathematical Problems in the Geophysical Sciences, No. 2: Inverse Problems, Dynamo Theory and Tides*, pp. 293–362. Providence, RI: American Mathematical Society.

Lindzen RS and Hou AY (1988) Hadley circulations for zonally averaged heating centred off the equator. *Journal of the Atmospheric Sciences* 45: 2416–2447.

Liou K-N (2002) *An Introduction to Atmospheric Radiation*, 2nd edn. Amsterdam, London: Academic Press.

Lodders K (2010) Atmospheric chemistry of the gas giant planets. *Geochemical News* 142: 1–11.

Lodders K and Fegley B (1998) *The Planetary Scientist's Companion*. New York: Oxford University Press.

López-Puertas M and Taylor FW (2001) *Non-LTE Radiative Transfer in the Atmosphere*. London: World Scientific.

Lorius C, Jouzel J, Raynaud D, Hansen J, and Letreut H (1990) The ice-core record: Climate sensitivity and future greenhouse warming. *Nature* 347: 139–145.

Lyons TW, Reinhard CT, and Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* 506: 307–315.

Magalhaes JA, Seiff A, and Young RE (2002) The stratification of Jupiter's troposphere at the Galileo probe entry site. *Icarus* 158: 410–433.

Mahaffy PR, Webster CR, Atreya SK, et al. (2013) Abundance and isotopic composition of gases in the martian atmosphere from the Curiosity rover. *Science* 341: 263–266.

Mak M (2011) *Atmospheric Dynamics*. Cambridge, NY: Cambridge University Press.

Manabe S and Strickler RF (1964) Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of the Atmospheric Sciences* 21: 361–385.

Manabe S and Wetherald RT (1967) Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences* 24: 241–259.

Marinova MM, McKay CP, and Hashimoto H (2005) Radiative-convective model of warming Mars with artificial greenhouse gases. *Journal of Geophysical Research* 110.

Martin LJ and Zurek RW (1993) An analysis of the history of dust activity on Mars. *Journal of Geophysical Research* 98: 3221–3246.

Martyn DF and Pulley OO (1936) The temperature and constituents of the upper atmosphere. *Proceedings of the Royal Society of London A* 154: 455–486.

McElroy MB and Donahue TM (1972) Stability of the Martian atmosphere. *Science* 177: 986–988.

McGouldrick K, Toon OB, and Grinspoon DH (2011) Sulfuric acid aerosols in the atmospheres of the terrestrial planets. *Planetary and Space Science* 59: 934–941.

McKay CP, Lorenz RD, and Lunine JI (1999) Analytic solutions for the antigreenhouse effect: Titan and the early Earth. *Icarus* 137: 56–61.

McKay CP, Pollack JB, and Courtin R (1989) The thermal structure of Titan's atmosphere. *Icarus* 80: 23–53.

McKay CP, Pollack JB, and Courtin R (1991) The greenhouse and antigreenhouse effects on Titan. *Science* 253: 1118–1121.

McNeall D, Halloran PR, Good P, and Betts RA (2011) Analyzing abrupt and nonlinear climate changes and their impacts. *WIREs Climate Change* 2: 663–686.
Meadows VS and Crisp D (1996) Ground-based near-infrared observations of the Venus nightside: The thermal structure and water abundance near the surface. *Journal of Geophysical Research* 100: 4595–4622.
Melosh HJ and Vickery AM (1989) Impact erosion of the primordial atmosphere of Mars. *Nature* 338: 487–489.
Menou K, Cho JYK, Seager S, and Hansen BMS (2003) "Weather" variability of close-in extrasolar giant planets. *Astrophysical Journal* 587: L113–L116.
Menou K and Rauscher E (2009) Atmospheric circulation of hot Jupiters: A shallow three-dimensional model. *Astrophysical Journal* 700: 887–897.
Michalski G, Bohlke JK, and Thiemens M (2004) Long term atmospheric deposition as the source of nitrate and other salts in the Atacama Desert, Chile: New evidence from mass-independent oxygen isotopic compositions. *Geochimica et Cosmochimica Acta* 68: 4023–4038.
Migliorini A, Grassi D, Montabone L, Lebonnois S, Drossart P, and Piccioni G (2012) Investigation of air temperature on the nightside of Venus derived from VIRTIS-H on board Venus-Express. *Icarus* 217: 640–647.
Mitchell JL and Vallis GK (2010) The transition to superrotation in terrestrial atmospheres. *Journal of Geophysical Research* 115.
Montmessin F, Bertaux JL, Lefevre F, et al. (2011) A layer of ozone detected in the nightside upper atmosphere of Venus. *Icarus* 216: 82–85.
Moroz VI, Ekonomov AP, Moshkin BE, et al. (1985) Solar and thermal radiation in the Venus atmosphere. *Advances in Space Research* 5: 197–232.
Moses JI, Fouchet T, Yelle RV, et al. (2004) The stratosphere of Jupiter. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites, and Magnetosphere*. New York: Cambridge University Press.
Moses JI, Visscher C, Fortney JJ, et al. (2011) Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b. *Astrophysical Journal* 737.
Moskalenko NI, Ilin YA, Parzhin CN, and Rodionov LV (1979) Pressure-induced infrared radiation absorption in atmospheres. *Bulletin of the Academy of Sciences of the USSR, Atmospheric and Oceanic Physics* 15: 632–637.
Nappo CJ (2002) *An Introduction to Atmospheric Gravity Waves*. London: Academic Press.
Navarra A and Boccaletti G (2002) Numerical general circulation experiments of sensitivity to Earth rotation rate. *Climate Dynamics* 19: 467–483.
Newman M and Leovy C (1992) Maintenance of strong rotational winds in Venus middle atmosphere by thermal tides. *Science* 257: 647–650.
Niemann HB, Atreya SK, Demick JE, et al. (2010) Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini–Huygens probe gas chromatograph mass spectrometer experiment. *Journal of Geophysical Research* 115.
Nutzman P, Charbonneau D, Winn JN, et al. (2009) A precise estimate of the radius of the exoplanet Hd 149026b from Spitzer photometry. *Astrophysical Journal* 692: 229–235.
Ortiz JL, Orton GS, Friedson AJ, Stewart ST, Fisher BM, and Spencer JR (1998) Evolution and persistence of 5-μm hot spots at the Galileo probe entry latitude. *Journal of Geophysical Research* 103: 23051–23069.
Parkinson TD and Hunten DM (1972) Spectroscopy and aeronomy of O_2 on Mars. *Journal of the Atmospheric Sciences* 29: 1380–1390.
Pavlov AA, Kasting JF, and Brown LL (2001) UV-shielding of NH_3 and O_2 by organic hazes in the Archean atmosphere. *Journal of Geophysical Research* 106: 23267–23287.
Pearl JC, Conrath BJ, Hanel RA, Pirraglia JA, and Coustenis A (1990) The albedo, effective temperature, and energy balance of Uranus, as determined from Voyager IRIS data. *Icarus* 84: 12–28.
Pechmann JB and Ingersoll AP (1984) Thermal tides in the atmosphere of Venus: Comparison of model results with observations. *Journal of the Atmospheric Sciences* 41: 3290–3313.
Peralta J, Hueso R, and Sanchez-Lavega A (2007) A reanalysis of Venus winds at two cloud levels from Galileo SSI images. *Icarus* 190: 469–477.
Peralta J, Luz D, Berry DL, et al. (2012) Solar migrating atmospheric tides in the winds of the polar region of Venus. *Icarus* 220: 958–970.
Petty GW (2006) *A First Course in Atmospheric Radiation*, 2nd edn. Madison, WI: Sundog Publishing.
Pierrehumbert RT (2010) *Principles of Planetary Climate*. Cambridge: Cambridge University Press.
Pollack JB (1991) Kuiper prize lecture: Present and past climates of the terrestrial planets. *Icarus* 91: 173–198.
Pollack JB, Colburn DS, Fraser M, Kahn R, Carlston CE, and Pidek D (1979) Properties and effects of dust particles suspended in the Martian atmosphere. *Journal of Geophysical Research* 84: 2929–2945.
Pollack JB, Toon OB, Ackerman TP, Mckay CP, and Turco RP (1983) Environmental effects of an impact-generated dust cloud: Implications for the cretaceous-tertiary extinctions. *Science* 219: 287–289.
Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active South Pole of Enceladus. *Science* 311: 1393–1401.
Prinn RG (2014) Ozone, hydroxyl radical, and oxidative capacity. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*. New York: Elsevier.
Ramanathan V and Carmichael G (2008) Global and regional climate changes due to black carbon. *Nature Geoscience* 1: 221–227.
Ramanathan V and Inamdar A (2006) The radiative forcing due to clouds and water vapor. In: Kiehl JT and Ramanathan V (eds.) *Frontiers of Climate Modeling*, pp. 119–151. New York: Cambridge University Press.
Ramirez RM, Kopparapu R, Zugger ME, Robinson TD, Freedman R, and Kasting JF (2014) Warming early Mars with CO_2 and H_2. *Nature Geoscience* 7: 59–63.
Read PL and Lewis SR (2003) *The Martian Climate Revisited: Atmosphere and Environment of a Desert Planet*. London: Praxis.
Rhines PB (1975) Waves and turbulence on a beta plane. *Journal of Fluid Mechanics* 69: 417–443.
Richardson MI and Wilson RJ (2002) A topographically forced asymmetry in the martian circulation and climate. *Nature* 416: 298–301.
Robinson TD and Catling DC (2012) An analytic radiative-convective model for planetary atmospheres. *Astrophysical Journal* 757: 104. http://dx.doi.org/10.1088/0004-637X/757/1/104.
Robinson TD and Catling DC (2014) Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency. *Nature Geoscience* 7: 12–15.
Robock A (2000) Volcanic eruptions and climate. *Reviews of Geophysics* 38: 191–219.
Robock A, Oman L, and Stenchikov GL (2007) Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences. *Journal of Geophysical Research* 112.
Rodgers CD and Walshaw CD (1966) Computation of infra-red cooling rate in planetary atmospheres. *Quarterly Journal of the Royal Meteorological Society* 92: 67–92.
Russell JM, Luo MZ, Cicerone RJ, and Deaver LE (1996) Satellite confirmation of the dominance of chlorofluorocarbons in the global stratospheric chlorine budget. *Nature* 379: 526–529.
Sagan C (1977) Reducing greenhouses and temperature history of Earth and Mars. *Nature* 269: 224–226.
Salby ML (2012) *Physics of the Atmosphere and Climate*, 2nd edn. New York: Cambridge University Press.
Salyk C, Ingersoll AP, Lorre J, Vasavada A, and Del Genio AD (2006) Interaction between eddies and mean flow in Jupiter's atmosphere: Analysis of Cassini imaging data. *Icarus* 185: 430–442.
Samuelson RE, Hanel RA, Kunde VG, and Maguire WC (1981) Mean molecular weight and hydrogen abundance of Titan's atmosphere. *Nature* 292: 688–693.
Sanchez-Lavega A (2010) *An Introduction to Planetary Atmospheres*. Boca Raton, FL: CRC Press/Taylor & Francis.
Sanchez-Lavega A, Perez-Hoyos S, and Hueso R (2004) Clouds in planetary atmospheres: A useful application of the Clausius–Clapeyron equation. *American Journal of Physics* 72: 767–774.
Satoh M (2004) *Atmospheric Circulation Dynamics and General Circulation Models*. Chichester, UK: Praxis Publishing, Springer-Praxis Books in Environmental Sciences.
Sausen R and Santer BD (2003) Use of changes in tropopause height to detect human influences on climate. *Meteorologische Zeitschrift* 12: 131–136.
Schmidt GA, Ruedy RA, Miller RL, and Lacis AA (2010) Attribution of the present-day total greenhouse effect. *Journal of Geophysical Research* 115: D20106. http://dx.doi.org/10.1029/2010JD014287.
Schmitt B, Debergh C, Lellouch E, Maillard JP, Barbe A, and Doute S (1994) Identification of 3 absorption bands in the 2 micron spectrum of Io. *Icarus* 111: 79–105.
Schneider T (2006) The general circulation of the atmosphere. *Annual Review of Earth and Planetary Sciences* 34: 655–688.
Schneider T and Liu JJ (2009) Formation of jets and equatorial superrotation on Jupiter. *Journal of the Atmospheric Sciences* 66: 579–601.
Schneider T and Walker CC (2006) Self-organization of atmospheric macroturbulence into critical states of weak nonlinear Eddy–Eddy interactions. *Journal of the Atmospheric Sciences* 63: 1569–1586.
Schofield JT and Taylor FW (1983) Measurements of the mean, solar-fixed temperature and cloud structure of the middle atmosphere of Venus. *Quarterly Journal of the Royal Meteorological Society* 109: 57–80.
Schubert G (1983) General circulation and the dynamical state of the Venus circulation. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 681–765. Tucson, AZ: University of Arizona Press.
Seager S (2010) *Exoplanet Atmospheres: Physical Processes*. Princeton, NJ: Princeton University Press.

Seager S and Deming D (2010) Exoplanet atmospheres. *Annual Review of Astronomy and Astrophysics* 48: 631–672.
Seager S, Dotson R, and Lunar and Planetary Institute (2010) *Exoplanets*, *The University of Arizona space science series*. Tucson, AZ: University of Arizona Press.
Seiff A, Kirk DB, Young RE, et al. (1980) Measurements of thermal structure and thermal contrasts in the atmosphere of Venus and related dynamical observations: Results from the four Pioneer Venus probes. *Journal of Geophysical Research* 85: 7903–7933.
Sekiya M, Hayashi C, and Nakazawa K (1981) Dissipation of the primordial terrestrial atmosphere due to irradiation of the solar far-UV during T-Tauri stage. *Progress in Theoretical Physics* 66: 1301–1316.
Sekiya M, Nakazawa K, and Hayashi C (1980) Dissipation of the primordial terrestrial atmosphere due to irradiation of the solar EUV. *Progress in Theoretical Physics* 64: 1968–1985.
Settle M (1979) Formation and deposition of volcanic sulfate aerosols on Mars. *Journal of Geophysical Research* 84: 8343–8354.
Shetty S and Marcus PS (2010) Changes in Jupiter's Great Red Spot (1979–2006) and Oval BA (2000–2006). *Icarus* 210: 182–201.
Shine KP, Barnett JJ, and Randel WJ (2008) Temperature trends derived from Stratospheric Sounding Unit Radiances: The effect of increasing CO_2 on the weighting function. *Geophysical Research Letters* 35.
Showman AP, Cho JYK, and Menou K (2010) Atmospheric circulation of extrasolar planets. In: Seager S (ed.) *Exoplanets*, pp. 471–516. Tucson, AZ: University of Arizona Press.
Showman AP, Wordsworth RD, Merlis TM, and Kaspi Y (2013) Atmospheric circulation of terrestrial exoplanets. In: Mackwell SJ, Simon-Miller AA, Harder JW, and Bullock MA (eds.) *Comparative Climatology of Terrestrial Planets*, pp. 277–326. Tucson, AZ: University of Arizona Press.
Smith ML, Claire MW, Catling DC, and Zahnle KJ (2014) The formation of sulfate, nitrate and perchlorate salts in the martian atmosphere. *Icarus* 231: 51–64.
Smith MD, Pearl JC, Conrath BJ, and Christensen PR (2001) Thermal Emission Spectrometer results: Mars atmospheric thermal structure and aerosol distribution. *Journal of Geophysical Research* 106: 23929–23945.
Smyth WH and Marconi ML (2006) Europa's atmosphere, gas tori, and magnetospheric implications. *Icarus* 181: 510–526.
Spencer JR, Jessup KL, McGrath MA, Ballester GE, and Yelle R (2000) Discovery of gaseous S_2 in Io's Pele plume. *Science* 288: 1208–1210.
Starr VP (1948) An essay on the general circulation of the Earth's atmosphere. *Journal of Meteorology* 5: 39–43.
Stephens GL and Tjemkes SA (1993) Water vapour and its role in the Earth's greenhouse. *Australian Journal of Physics* 46: 149–166.
Taylor JR (2005) *Classical Mechanics*. Sausalito, CA: University Science Books.
Taylor FW (2010) *Planetary Atmospheres*. Oxford, NY: Oxford University Press.
Tenishev V, Combi MR, Teolis BD, and Waite JH (2010) An approach to numerical simulation of the gas distribution in the atmosphere of Enceladus. *Journal of Geophysical Research* 115.
Teolis BD, Jones GH, Miles PF, et al. (2010) Cassini finds an oxygen-carbon dioxide atmosphere at Saturn's icy moon Rhea. *Science* 330: 1813–1815.
Thomas GE and Stamnes K (1999) *Radiative Transfer in the Atmosphere and Ocean*. Cambridge, NY: Cambridge University Press, Cambridge Atmospheric and Space Science Series.
Thompson RORY (1971) Why there is an intense eastward current in the North Atlantic but not in the South Atlantic. *Journal of Physical Oceanography* 1: 235–237.
Thompson DWJ, Solomon S, Kushner PJ, England MH, Grise KM, and Karoly DJ (2011) Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature Geoscience* 4: 741–749.
Tian F, Claire MW, Haqq-Misra JD, et al. (2010) Photochemical and climate consequences of sulfur outgassing on early Mars. *Earth and Planetary Science Letters* 295: 412–418.
Titov DV, Bullock MA, Crisp D, Renno NO, Taylor FW, and Zasova LV (2007) Radiation in the atmosphere of Venus. In: Esposito LW, Stofan ER, and Cravens TE (eds.) *Exploring Venus as a Terrestrial Planet*, pp. 121–138. Washington, D.C.: AGU.
Tokano T (2009) The dynamics of Titan's troposphere. *Philosophical Transactions of the Royal Society* A367: 633–648.
Tokano T (2010) Relevance of fast westerlies at equinox for the eastward elongation of Titan's dunes. *Aeolian Research* 2: 113–127.
Tokar RL, Johnson RE, Thomsen MF, et al. (2012) Detection of exospheric O_2^+ at Saturn's moon Dione. *Geophysical Research Letters* 39.
Tomasko MG and West RA (2009) Aerosols in Titan's atmosphere. In: Brown RH, Lebreton J-P, and Waite JH (eds.) *Titan from Cassini–Huygens*, pp. 297–322. New York: Springer.
Trafton LM (1966) Pressure-induced monochromatic translational absorption coefficients for homopolar and non-polar gases and gas mixtures with particular application to H_2. *Astrophysical Journal* 146, p. 558.
Trafton LM (1998) Planetary atmospheres: The role of collision-induced absorption. In: Vigasin AA and Slanina Z (eds.) *Molecular Complexes in Earth's Planetary, Cometary and Interstellar Atmospheres*, pp. 177–193. London: World Scientific.
Trafton LM, Hunten DM, Zahnle KJ, and McNutt RL (1997) Escape processes at Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 475–522. Tucson, AZ: University of Arizona Press.
Trenberth KE, Fasullo JT, and Kiehl J (2009) Earth's global energy budget. *Bulletin of the American Meteorological Society* 90: 311–324.
Turco RP, Toon OB, Ackerman TP, Pollack JB, and Sagan C (1983) Nuclear winter: Global consequences of multiple nuclear explosions. *Science* 222: 1293–1300.
Vallis GK (2006) *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale circulation*. New York: Cambridge University Press.
Vardavas IM and Taylor FW (2007) *Radiation and Climate*. Oxford: Oxford University Press, International Series of Monographs on Physics.
Vasavada AR and Showman AP (2005) Jovian atmospheric dynamics: An update after Galileo and Cassini. *Reports on Progress in Physics* 68: 1935–1996.
Walker JCG (1977) *Evolution of the Atmosphere*. New York: Macmillan.
Walker JCG (1982) The earliest atmosphere of the Earth. *Precambrian Research* 17: 147–171.
Walker JCG (1986) Impact erosion of planetary atmospheres. *Icarus* 68: 87–98.
Walker JCG, Hays PB, and Kasting JF (1981) A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86: 9776–9782.
Wallace JM and Hobbs PV (2006) *Atmospheric Science: An Introductory Survey*, 2nd edn. Burlington, MA: Elsevier Academic Press, International geophysics series.
Watson AJ, Donahue TM, and Walker JCG (1981) The dynamics of a rapidly escaping atmosphere: applications to the evolution of Earth and Venus. *Icarus* 48: 150–166.
Weaver CP and Ramanathan V (1995) Deductions from a simple climate model: Factors governing surface temperature and atmospheric thermal structure. *Journal of Geophysical Research* 100: 11585–11591.
Webster CR, Mahaffy PR, Atreya SK, Flesch GJ, Farley KA, and Team MS (2013) Low upper limit to methane abundance on Mars. *Science* 342: 355–357.
West RA, Baines KH, Friedson AJ, Banfield D, Ragent B, and Taylor FW (2004) Jovian clouds and haze. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites, and Magnetosphere*, pp. 79–104. New York: Cambridge University Press.
West RA, Baines KH, Karkoschka E, and Sanchez-Lavega A (2009) Clouds and aerosols in Saturn's atmosphere. In: Dougherty M, Esposito L, and Krimigis S (eds.) *Saturn from Cassini–Huygens*, pp. 161–180. New York: Springer.
West RA, Baines KH, and Pollack JB (1991) Clouds and aerosols in the Uranian atmosphere. In: Bergstralh JT, Miner ED, and Matthews MS (eds.) *Uranus*, pp. 296–324. Tucson, AZ: University of Arizona Press.
Williams GP (1978) Planetary circulations.1. Barotropic representation of Jovian and terrestrial turbulence. *Journal of the Atmospheric Sciences* 35: 1399–1426.
Wilson EH and Atreya SK (2004) Current state of modeling the photochemistry of Titan's mutually dependent atmosphere and ionosphere. *Journal of Geophysical Research* 109: E06002. http://dx.doi.org/10.1029/2003JE002181.
Wilson RJ, Banfield D, Conrath BJ, and Smith MD (2002) Traveling waves in the Northern Hemisphere of Mars. *Geophysical Research Letters* 29.
Wolf ET and Toon OB (2010) Fractal organic hazes provided an ultraviolet shield for early Earth. *Science* 328: 1266–1268.
Wordsworth R, Forget F, and Eymet V (2010) Infrared collision-induced and far-line absorption in dense CO_2 atmospheres. *Icarus* 210: 992–997.
Wordsworth R and Pierrehumbert R (2013) Hydrogen-nitrogen greenhouse warming in Earth's early atmosphere. *Science* 339: 64–67.
Yamamoto M and Takahashi M (2009) Dynamical effects of solar heating below the cloud layer in a Venus-like atmosphere. *Journal of Geophysical Research* 114.
Yang J, Gong P, Fu R, et al. (2013) The role of satellite remote sensing in climate change studies. *Nature Climate Change* 3: 875–883.
Yelle RV, Griffith CA, and Young LA (2001) Structure of the Jovian stratosphere at the Galileo probe entry site. *Icarus* 152: 331–346.
Yung YL, Allen M, and Pinto JP (1984) Photochemistry of the atmosphere of Titan: Comparison between model and observations. *The Astrophysical Journal, Supplement Series* 55: 465–506.
Yung YL and DeMore WB (1999) *Photochemistry of Planetary Atmospheres*. New York: Oxford University Press.
Zahnle K, Freedman RS, and Catling DC (2011) Is there methane on Mars? *Icarus* 212: 493–503.

Zahnle K, Haberle RM, Catling DC, and Kasting JF (2008) Photochemical instability of the ancient Martian atmosphere. *Journal of Geophysal Research* 113: E11004.

Zahnle KJ and Kasting JF (1986) Mass fractionation during transonic escape and implications for loss of water from Mars and Venus. *Icarus* 68: 462–480.

Zahnle K, Kasting JF, and Pollack JB (1990) Mass fractionation of noble gases in diffusion-limited hydrodynamic hydrogen escape. *Icarus* 84: 502–527.

Zahnle K, Pollack JB, Grinspoon D, and Dones L (1992) Impact-generated atmospheres over Titan, Ganymede, and Callisto. *Icarus* 95: 1–23.

Zalucha AM, Gulbis AAS, Zhu X, Strobel DF, and Elliot JL (2011) An analysis of Pluto occultation light curves using an atmospheric radiative-conductive model. *Icarus* 211: 804–818.

Zelinka MD and Hartmann DL (2011) The observed sensitivity of high clouds to mean surface temperature anomalies in the tropics. *Journal of Geophysical Research* 116.

Zerkle AL, Claire M, Domagal-Goldman SD, Farquhar J, and Poulton SW (2012) A bistable organic-rich atmosphere on the Neoarchaean Earth. *Nature Geoscience* 5: 359–363.

Zhang X, Nixon CA, Shia RL, et al. (2013) Radiative forcing of the stratosphere of Jupiter, Part I: Atmospheric cooling rates from Voyager to Cassini. *Planetary and Space Science* 88: 3–25.

10.14 Geology, Life, and Habitability

G Southam, The University of Queensland, Brisbane, QLD, Australia
F Westall, Centre de Biophysique Moléculaire, Orléans, France
T Spohn, German Aerospace Center (DLR), Berlin, Germany

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by G Southam and F Westall, Volume 10, pp. 421–437, © 2007, Elsevier B.V.

10.14.1	**Introduction**	473
10.14.1.1	The Requirements for an Origin of Life and 'Habitability'	473
10.14.1.2	The Origin of Life and the First Cells	474
10.14.2	**Geology, Life, and Habitability**	474
10.14.2.1	Habitability as Defined by the Physical and Chemical Environmental Constraints on Life	475
10.14.2.2	Building Cells	476
10.14.2.3	Generating Energy	477
10.14.2.4	The Growth of Microorganisms	478
10.14.3	**Geology and Life**	479
10.14.3.1	Lithotrophy	479
10.14.3.2	Microorganisms and Mineral Formation	480
10.14.3.3	Habitable Conditions on Early Earth (and Other Rocky Planets)	480
10.14.3.4	The Earliest Traces of Life	481
10.14.4	**Conclusions for Geology, Life, and Habitability Beyond the Earth**	482
References		483

10.14.1 Introduction

10.14.1.1 The Requirements for an Origin of Life and 'Habitability'

Whether or not life originated on Earth, it was certainly present in abundance and at a relatively advanced level of evolution by about 3.5 billion years ago (Ga) (see reviews in Westall and Southam (2006) and Westall (2011)). At this time, the habitable environment of the Earth was very much different than the Earth of today – significantly hotter (Arndt, 1994; Knauth and Lowe, 2003) and more volcanically/hydrothermally active (Hofmann and Bolhar, 2007) and basically without oxygen (<0.2% present atmospheric levels, Catling and Claire, 2005; Kasting, 1993). The low levels of reactive oxygen produced by photolysis (Kasting, 1993) relative to the abundance of reduced chemical species would have resulted in a correspondingly reducing chemistry for the hydrosphere and lithosphere as well. Based on comparisons with the lunar cratering record, the environment was also affected by a short, sharp influx of bolides (the Late Heavy Bombardment (LHB); Gomes et al., 2005) about 3.9 Ga ago that may or may not have evaporated all the surface water (Sleep et al., 1989) and eliminated all life (Maher and Stevenson, 1988). It is commonly believed that the life on Earth today evolved from primitive cells that could only have appeared after the end of the LHB (Lowe and Byerly, 1986; Lowe et al., 2003), that is, after ~3.85 Ga. However, the last decade has seen an enormous advance in our understanding of the limits of life and it is clear that, once started, it is very difficult to eradicate prokaryotes from either the Bacteria or Archaea domains (Woese et al., 1990), especially the latter since they generally occur in environmental conditions that today are considered to be extreme but were common on the early Earth. It is possible, therefore, that life appeared on Earth before ~4.0 Ga and survived the catastrophic LHB in, for instance, a protected subsurface habitat (cf. Russell et al., 2010), to reappear and flourish at the surface once water had recondensed. They may represent the survivors of a 'hot bottleneck' through which the early forms of life had to pass, such as would have existed during a catastrophic impact event. However, whether or not they represent the last universal common ancestor (LUCA) is still debated (Brochier and Philippe, 2002). An alternative hypothesis is that life first appeared on another planet and was transported by panspermia to the Earth (e.g., Cockell, 2008) once habitable conditions had been reestablished at the end of the catastrophic bolide period. However, the other likely contender for an independent origin of life in the inner part of the solar system, Mars, also suffered from the same bolide bombardment.

In order to address the concept of life and habitability in the solar system, we start with a brief overview of the current thinking on the basic requirements for life to originate and the origin of life. In life as we know it, all cells need a source of energy and are composed of water and carbon molecules that include some essential elements (notably hydrogen, oxygen, nitrogen, phosphorus, and sulfur (HONPS)). These are therefore the minimum requirements on a planet (or other type of rocky body) for the appearance of life. Since organic carbon molecules occur in interstellar clouds, as well as in the materials forming the proto-solar system cloud (in the form of the planetesimals that formed the rocky planets, asteroids, meteorites, micrometeorites, and comets), they can be considered to be more or less ubiquitous. Endogenous sources of organic carbon included the primordial, slightly reducing atmosphere (Johnson et al., 2008; Miller, 1953) and active hydrothermal systems where organic carbon is primarily produced via Fischer–Tropsch synthesis (i.e., hydrocarbons containing up

to 29 carbon atoms have been produced from the Rainbow ultramafic hydrothermal system on the Mid-Atlantic Ridge; Holm and Charlou, 2001; see also Russell et al., 2010), which are believed to have been very active on the early Earth. A significant amount of the organic molecules (and other volatiles, such as water) were also delivered from extraterrestrial sources (Chyba and Sagan, 1992; Sephton and Botta, 2008). Carbonaceous chondrite meteorites contain up to 5% organic carbon, and eight of the amino acids that have been identified in them are of the kind used in living cells today; they are the building blocks of life. The present-day flux of carbonaceous micrometeorites (to the Earth) containing up to 2% carbon is estimated between 50 and 100 tons per year, which implies an extraordinary amount equaling 10^{23} g carbon over a 300 My period during the LHB (Maurette et al., 2000; see also Brack, 2004, for a review). Thus, organic carbon on rocky bodies in the young solar system was readily available.

The sources of the other elements essential for an origin of life (HONPS) would likewise be readily available on any rocky planet. A rocky planet is also an excellent source of energy. Living organisms need energy in order to militate against the constant problem of entropy, the increase in disorder or dissipation of chemical energy within a closed system, their entire lives. From a simplistic, energy perspective then, any geologic system that provides geochemical energy, that is, a reaction that can be written possessing a negative ΔG_r, should conceivably have the capacity to support life. However, because water is a necessary ingredient of life as we know it and is therefore also necessary for life to emerge elsewhere in our solar system (or in the universe), this means that the target planetary body should be situated within the habitable zone around its star (see Franck et al., 2006; Kasting et al., 1993; Kopparapu et al., 2013). Habitable conditions are defined by the sum of the physical and chemical conditions, which support the presence of liquid water at the surface of a planetary body. This habitable zone around the Sun has moved outward since the formation of the solar system as luminosity, now about 25–30% >4.6 Ga ago, has increased (Newman and Rood, 1977). While rocky planets within this zone represent ideal potential habitats, rocky bodies with liquid water in contact with rock do exist outside this zone and can thus be considered as being potentially habitable. In the early part of the history of the solar system, a number of planets and satellites had their own internal heat source from decaying radioactive elements, and in the case of Europa, there is convincing evidence that a liquid ocean exists underneath its icy crust (e.g., Hussmann et al., 2014; **Chapter 10.18**) with the energy required to keep the fluid liquid being impacted by the strong tidal forces from Jupiter (Chyba and Phillips, 2001; Hussmann et al. 2014) and the Laplace orbital resonance of the three inner Galilean satellites Io, Europa, and Ganymede. Analysis of the induced magnetic field (Khurana et al., 1998) as well as more recent spectral observation suggests that Europa's ocean is a brine containing Na, K, and Mg salts (Brown and Hand, 2013).

10.14.1.2 The Origin of Life and the First Cells

A working definition of a minimal life-form could be "a structure capable of auto-reproduction including errors that lead to evolution" (Brack, 2005). The basic structure consists of three long chain carbon molecules that are necessary for the functioning of the cell: (1) a lipid membrane enclosing a (2) molecule capable of information transfer (RNA/DNA) and (3) a catalyzing molecule (enzyme) to provide the basic work of the cell. Despite the fact that it has not yet been possible to reproduce a life-form *ex novo* in the laboratory, the last 50 years has seen significant advances in understanding how these basic elements formed from inorganic components.

In the modern cell, membranes are primarily composed of amphipathic phospholipids that are stable in water. These lipids can be formed from fatty acids found in carbonaceous chondrites like the Murchison meteorite (Deamer, 1998). Catalytic reactions in modern, living cells are performed by proteinaceous enzymes, themselves made of homochiral amino acids. As we have noted earlier, these were abundantly available on the early Earth. There are, however, a number of scenarios regarding the condensation or assemblage of simple amino acids into longer chain molecules, such as peptides. These include condensation on mineral surfaces, such as clays (Ferris et al., 1996), thermal condensation (Fox and Dose, 1977), and chemical reactions (Barbier and Brack, 2002). Likewise, the origin of the first information transfer molecules is highly debated. Nucleic acids in modern cells are composed of bases (purine and pyrimidine), sugars (ribose or deoxyribose), and phosphate groups. RNA (ribonucleic acid) is considered to be a likely, primitive replicating molecule because it also has catalytic properties (representing self-replication, mutation, and evolution) and can act as a polymerization template (Zaug and Cech, 1986). However, even RNA is too complicated to have been the first replicating molecule (Schwartz and Orgel, 1985). Nonenzymatic replication has been demonstrated (Inoue and Orgel, 1982), as have autocatalytic reactions involving the reaction of FeS and H_2S on the surface of pyrite (Wächtershäuser, 1988). In the latter scenario, organic molecules forming a network on the surface of the mineral evolve. Other candidates for precursor replication molecules include pyranosyl-RNA and threo furanosyl nucleic acid, though the latter is considered to be more likely because of the greater availability of the carbon precursors of tetrose than ribose on the primitive Earth.

The basic process leading to an origin of life can be envisaged thus (Brack, 2004):

$$\text{CHONPS} + \text{water} \rightarrow \text{catalyzers} \rightarrow \text{RNA world} \rightarrow \text{cells (RNA, proteins, and membranes)}$$

10.14.2 Geology, Life, and Habitability

Based on the previously described criteria for habitability and the origin of life, the key to understanding the potential for the emergence of life, that is, habitability, is the understanding of the history of liquid water on any terrestrial planet or icy moon. Given the necessity of disequilibria conditions in the emergence of life, liquid water needs to be in contact with minerals/rocks, implying liquid water on a rocky planet. The capacity to support life may be transient, that is, it will not be constant for a particular planet over time but will vary based on a wide range of geologic and planetary factors (e.g., Westall et al., 2013). We only have terrestrial life to use as a basis for

discussion of the relationship between geology (including aqueous geochemistry), life, and habitability. Assuming water is available, there is a wide diversity of metabolic strategies that have been developed by terrestrial microorganisms (Barns and Nierzwicki-Bauer, 1997; Pace, 1997; Reysenbach and Shock, 2002), so that they can successfully colonize almost all geologic environments.

In this chapter, we focus on the interaction between geology and prokaryotic life as a way to address the subject of habitability, both from our perspective and from that of the prokaryotes, which have inhabited Earth for far longer than we have. Today, prokaryotic organisms are ubiquitous on Earth, being found in a wide variety of surface and subsurface environments (Colwell et al., 1997; Lovley and Chapelle, 1995; Onstott et al., 2006; Parkes et al., 1994; Stevens et al., 1993), where they are only limited by the availability of water (Navarro-Gonzalez et al., 2003) and temperatures $<121\,^{\circ}C$ (Kashefi and Lovley, 2003). In these natural environments, prokaryotes typically grow as biofilms on mineral surfaces (Marshall, 1988; Wanger et al., 2006). This intimate association of life with its geologic substrate is thought to have occurred for at least the last ~3.5 billion years (see Westall and Southam, 2006; Westall, 2011, for a review) producing molecular to global-scale by-products. During the initial stages of this interaction, the physical and chemical conditions of the early Earth would have controlled the biosphere. However, over time, life exerted an increasing influence over the geosphere. Life serves as a catalyst for a wide array of reactions at geologically low temperatures, doing them faster than would happen abiotically (e.g., Singer and Stumm, 1970).

10.14.2.1 Habitability as Defined by the Physical and Chemical Environmental Constraints on Life

The growth of prokaryotes is affected by the various physical and chemical constraints of the environments in which they are growing (Figure 1). They, in turn, can have a profound effect on their surrounding environment. Understanding the physical and chemical limits of bacteria and their environmental influences allows us to predict the types and distribution of microorganisms that could occur in natural systems.

The range of (and optimum) growth temperatures differentiate prokaryotes into broad groups (Stetter et al., 1990). Psychrophiles grow optimally at or below 15 °C and have even demonstrated activity down to −20 °C (Mykytczuk et al., 2013); viable prokaryotes have been recovered from million-year-old permafrost (Gilichinsky et al., 2007). Psychrotolerant prokaryotes grow optimally between 20 and 40 °C but will grow at temperatures as low as 0 °C. Mesophiles grow optimally between 15 and 45 °C. Thermophiles have optimum growth temperatures above 45 °C and hyperthermophiles grow optimally above 80 °C (Figure 2). While no single microorganism can withstand the full range of temperatures described here, as a group, prokaryotes can be found across any thermal gradient, from ice or permafrost regions all the way up to hydrothermal fluids (121°C).

In all environments, prokaryotes concentrate solutes within their cell envelopes and obtain water through osmosis, thus enabling them to maintain intra-cell turgor pressure (Kunte, 2006). The prokaryote cell envelope is designed to withstand this turgor pressure, which can measure between 2-3 atmospheres in gram-negative bacteria and up to 15 atmospheres in gram-positive bacteria (Beveridge, 1981). The accumulation of water represents a challenge for prokaryotes in the Archaea domain in contemporary (e.g., the Great Salt Lake) and paleo-evaporitic (Bechtel et al., 1996), hypersaline systems. Some of these environments are continually under stress due to low water activity and are therefore considered to be examples of extreme environments, which will only support halophilic (or halotolerant) microorganisms (see Figure 3).

Most natural environments occur as circumneutral pH systems. Therefore, most prokaryotes exhibit optimal activity in the pH range of 6–8. At pH 7, the hydrogen ion concentration is 100 nM. Extreme environments possessing acidic (e.g., pH 2) or alkaline (e.g., pH 11) pH conditions will contain

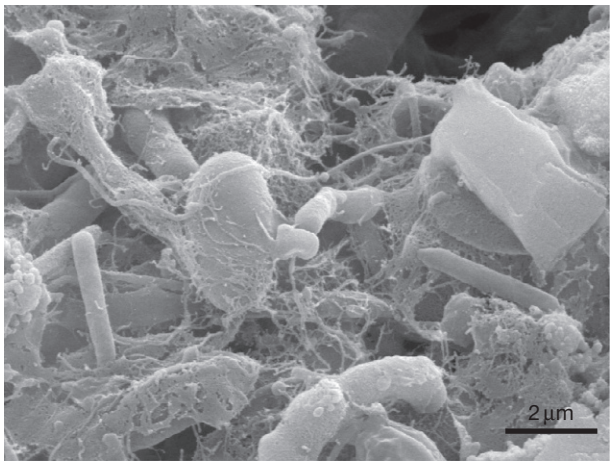


Figure 1 A scanning electron micrograph of a bacterial biofilm collected from a groundwater outflow into an oxygenated mining environment, 3.1 km below land surface in the Witwatersrand Basin, Republic of South Africa. Note the spiderweb-like appearance of the bacterial exopolymer material.

Figure 2 A photograph of a hydrothermal spring in Norris Geyser Basin, Yellowstone National Park. Extreme environments, such as this, are dominated by prokaryotic organisms representing both the Archaea and Bacteria domains. The field of view is ~8 m across.

Figure 3 A photograph of Basque Lake #2, in central BC, Canada. The evaporite minerals in this basin are dominated by epsomite, indicating the hypersaline nature of this environment.

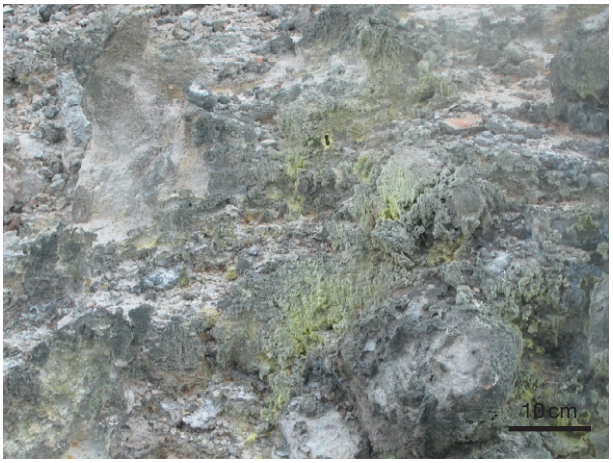


Figure 4 A photograph of a hydrothermal acid hot spring in Shimabara, Japan. In these extreme environments, the geosphere dominates the biosphere.

indigenous microorganisms, that is, acidophilic (acidotolerant) or alkaliphilic (alkalitolerant) prokaryotes, respectively, whose growth parameters are adapted to these extremes in environmental pH (Goodwin and Zeikus, 1987). Note that the word 'extremophile' is an anthropocentric term because for many microorganisms, thiobacilli, an ambient pH of 2 is optimum.

Our modern habitat is characterized by an atmosphere rich in oxygen (which was not the case on the early Earth). The capacity to utilize oxygen as a terminal electron acceptor creates a competitive advantage as well as a challenge for prokaryotes, which are subdivided into aerobic and anaerobic groups based on their oxygen requirements and detoxification strategies (Chapelle, 1993). There are also facultatively anaerobic prokaryotes that utilize oxygen if it is available, becoming anaerobes when the oxygen runs out, that is, they can grow with or without oxygen. The utilization of oxygen as a terminal electron acceptor enables organisms to fully oxidize their respective source(s) of energy and thereby maximize their energy gain from reaction [1]:

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O;\ \Delta G^o = -2870\,kJ\,mol^{-1} \quad [1]$$

Any organism that is capable of growing in the presence of oxygen, whether or not it utilizes oxygen as a terminal electron acceptor, must also have the capacity to detoxify the chemically reactive oxygen compounds produced through the stepwise reduction of oxygen to water (hydrogen peroxide, superoxide radical, and the hydroxyl radical; Atlas and Bartha, 1997). Detoxification of these oxygen radicals requires a series of enzymes including superoxide dismutase, catalase, and peroxidase. The ability to tolerate oxygen relates to the presence or absence of each of these enzymes. Measurements of dissolved oxygen and redox potential, which represents the sum reactivity of oxidants and reductants, are important in determining which microbial processes are functioning within natural systems.

Analyses of the physical and chemical limits on prokaryote growth and survival typically focus on one 'stressor' at a time, for example, pH (acidophile vs. neutrophile vs. alkaliphile), temperature (psychrophile vs. mesophile vs. thermophile), salinity (freshwater vs. halophile), or oxygen availability (aerobe vs. anaerobe). On the other hand, there are many examples of prokaryotes that can withstand multiple 'stressors' (see de Vera et al., 2013), for example, thermoacidophiles. It is in these environments (see **Figure 4**) that the diversity of prokaryotes and the limits of life are truly appreciated (see **Table 1**). Nevertheless, other combinations of 'stressors' will be deadly to microorganisms, such as low water activity together with high UV radiation, as on the surface of Mars today (Cockell and Raven, 2004).

10.14.2.2 Building Cells

Modern prokaryotes use a wide range of organic and inorganic nutrients. Organic nutrients, for example, carbohydrates, amino acids, nucleic acids, and hydrocarbons, include nearly every organic monomer or polymer found in the living or diagenetically altered biosphere. Inorganic nutrients include dissolved gases (carbon monoxide, carbon dioxide, hydrogen, nitrogen, and dihydrogen sulfide), soluble cations (sodium, calcium, magnesium, ammonium, and ferrous and ferric iron), base metals (chromium, nickel, copper, cobalt, zinc, lead, and mercury), and soluble anions (chloride, nitrite, nitrate, hydrogen sulfide, sulfite, sulfate, phosphate, selenate, and arsenate). Depending on nutrient requirements, these compounds can be used in assimilation reactions, that is, building new cells, or in dissimilatory processes to generate the energy needed to construct new biomass (discussed in the succeeding text).

Carbon utilization is one of the most important criteria by which prokaryotes are characterized. Heterotrophic prokaryotes require organic carbon for their metabolism and typically couple the oxidation of organic carbon to CO_2 with the reduction of dissolved inorganic species or minerals. The sources of carbon and energy for heterotrophs can range from simple organics such as glucose and asparagine, an amino acid (Goldman and Wilson, 1977), to macromolecular materials such as cellulose (Ljungdahl and Eriksson, 1985).

Table 1 Classification and examples of extremophiles

Environmental parameter	*Type*	*Definition*	*Examples*
Temperature	Hyperthermophile	Growth >80 °C	*Pyrolobus fumarii*, 113 °C
	Thermophile	Growth 60–80 °C	*Synechococcus lividis*
	Mesophile	15–60 °C	*Homo sapiens*
	Psychrophile	<15 °C	*Psychrobacter*, some insects
Radiation			*Deinococcus radiodurans*
Pressure	Barophile	Weight loving	Unknown
	Piezophile	Pressure loving	For microbe, 130 MPa
Gravity	Hypergravity	>1 g	None known
	Hypogravity	<1 g	None known
Vacuum		Tolerates vacuum (space devoid of matter)	Tardigrades, insects, microbes, seeds
Desiccation	Xerophiles	Anhydrobiotic	*Artemia salina*; nematodes, microbes, fungi, lichens
Salinity	Halophile	Salt loving (2–5 M NaCl)	*Halobacteriacea*, *Dunaliella salina*
pH	Alkaliphile	pH > 9	*Natronobacterium*, *Bacillus firmus* OF4, *Spirulina spp.* (all pH 10.5)
	Acidophile	Low pH loving	*Cyanidium caldarium*, *Ferroplasma sp.* (both pH 0)
Oxygen tension	Anaerobe	Cannot tolerate O_2	*Methanococcus jannaschii*
	microaerophile	Tolerates some O_2	*Clostridium*
	aerobe	requires O_2	*Homo sapiens*
Chemical extremes	Gases metals	Can tolerate high concentrations of metal (metalotolerant)	*Cyanidium caldarium* (pure CO_2) *Ferroplasma acidarmanus* (Cu, As, Cd, Zn); *Ralstonia sp.* CH34 (Zn, Co, Cd, Hg, Pb)

Reproduced from Rothschild and Mancinelli (2001).

Autotrophs are able to obtain their cellular carbon for biomass from inorganic sources (dissolved CO_2/HCO_3^-). Autotrophic organisms direct most of their energy toward the fixation (reduction) of CO_2 into organic carbon, that is, biomass production, and therefore tend to grow at slower rates than heterotrophic organisms. However, autotrophs are extremely important since they serve as the basis of the food chain upon which 'we' heterotrophs thrive. For example, autotrophic bacteria oxidize sulfur (reaction [2]) and synthesize carbohydrates at seafloor hot springs (reaction [3]):

$$2H_2S + O_2 \rightarrow 2S^\circ + 2H_2O;\ \Delta G^o = -441\,kJ\,mol^{-1} \quad [2]$$

$$CO_2 + 2H_2S \rightarrow CH_2O + 2S^\circ + H_2O \quad [3]$$

They are the base of the 'food chain' in these lightless, extreme environments (Cavanaugh et al., 1981).

Nitrogen and phosphorus are the limiting nutrients in most natural systems. However, prokaryotes can easily assimilate all forms of water-soluble inorganic nitrogen, and when the availability of nitrogen is limited, a wide range of prokaryotes possess a nitrogenase enzyme that allows them to fix atmospheric nitrogen (Brill, 1975). Regarding phosphorus-limiting conditions, microorganisms can produce a wide array of organic and inorganic acids that can enhance the weathering of silicate minerals in rocks and thus assimilate phosphate (Rogers et al., 1998; discussed in **Section 10.14.3.1**).

10.14.2.3 Generating Energy

In natural systems, prokaryotes exploit a wide range of redox reactions having negative ΔG_r to support metabolism (Nealson and Stahl, 1997). The requirement of energy for biosynthetic reactions is described according to the electron donor or source of reducing power and by the electron acceptor, that is, a redox couple. Current research on the diversity of life focuses on phenotypic and genotypic studies of geochemical gradients, where thermodynamic calculations highlight possible biogeochemical reactions that could support life, for example, anaerobic methane oxidation at deep-sea cold clathrate vents (Thomsen et al., 2001).

In natural systems, microbially catalyzed redox processes are established through competition for available nutrients (assimilation) and through the efficiency of their respective energy generation mechanisms (Lovley and Klug, 1983). As geochemical conditions become increasingly reducing, the amount of energy (as expressed by ΔG_r) available from each of the predominant biogeochemical reactions (i.e., redox zones) decreases (Lovley and Chapelle, 1995; Lovley and Phillips, 1988) resulting in a concomitant reduction in biomass (Onstott et al., 2006).

Phototrophs utilize sunlight to produce ATP (adenosine triphosphate, the principal energy carrier of all cells) from otherwise chemically stable reduced compounds. Anoxygenic phototrophs were probably already present on the Earth by 3.5 Ga (see reviews in Westall and Southam, 2006; Westall, 2011); contemporary representatives of this group are anaerobic green and purple sulfur bacteria that derive their energy for growth from light and H_2S or Fe^{2+} (Heising and Schink, 1998; Kappler and Newman, 2004), producing elemental sulfur, sulfate, or Fe^{3+} as end products of their metabolism (van Gemerden, 1986). The later-evolved oxygenic phototrophs use light in combination with H_2O as their source of energy (reaction [4]):

$$6CO_2 + 6H_2O + \text{sunlight} \rightarrow C_6H_{12}O_6 + 6O_2 \quad [4]$$

Chemolithotrophs (hydrogen, ammonia, or sulfur oxidizers and methylotrophs) are prokaryotes that utilize inorganic chemicals (H_2, NH_4^+, H_2S, and CH_4) as their energy

sources (electron donors). Lithotrophs grow more slowly than heterotrophs owing to the relatively low levels of chemical energy present in inorganic versus organic compounds (Nealson and Stahl, 1997) (**Figure 5**). Chemoorganotrophs, which utilize organic carbon as their source of reducing power and cellular carbon, include both aerobic and anaerobic prokaryotes.

Prior to the evolution of oxygenic photosynthesis (Dismukes et al., 2001; Kasting, 1993), the extremely low levels of oxygen in the atmosphere (<0.2% of present atmospheric levels, Kasting, 1993) would have resulted in lower overall biological activity (Des Marais, 2000). The capacity of oxygen to help provide abundant energy directly via aerobic respiration would simply not have existed. The small amount of oxygen produced by photolysis of H_2O vapor in the upper atmosphere or by radiolysis of water molecules at the surface of the oceans would have been immediately scavenged by the abundant reduced mineral species that covered the Earth, thus indirectly providing some oxidized inorganic electron acceptors, although nowhere near the quantity that was available ever since the atmosphere was oxidized.

Today, because of the presence of oxygen in our atmosphere and dissolved oxygen in surface and near-surface groundwater (**Figure 6**), aerobic microorganisms dominate the Earth's surface with anaerobic prokaryotes becoming increasingly important with depth below the surface. During the early to mid-Archean, prior to the evolution of oxygenic photosynthetic bacteria, or presumably, on any geochemically reducing planetary body possessing an 'early stage' biosphere, anaerobic organisms would have been the dominant group (Kasting, 1993).

Below the aerobic zone (or at the surface of an anaerobic planet), a progression of different anaerobes (e.g., metal reducers, sulfate reducers, and methanogens) exists (would exist), most of which (from Earth's lessons) utilize organic matter for their metabolism (Lovley and Chapelle, 1995; Lovley and Klug, 1983). Under moderately reducing conditions, Fe- and Mn-reducing bacteria couple the oxidation of organic matter to the reductive dissolution of high-surface-area Fe and Mn oxyhydroxides (Lovley and Phillips, 1988; Myers and Nealson, 1988). At increasingly reduced conditions, sulfate-reducing bacteria oxidize low-molecular-weight organic compounds, utilizing sulfate as the terminal electron acceptor, forming and releasing hydrogen sulfide as a by-product of metabolism (Anderson et al., 1998; Donald and Southam, 1999). Bacterial methane production occurs under even more reducing conditions than microbial sulfate reduction. Methanogenesis (reaction [5]) is also the dominant anaerobic prokaryotic process that cycles inorganic carbon back into the organic carbon pool (reaction [6]):

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O,\ \Delta G^o = -130.2\,kJ\,mol^{-1} \quad [5]$$

$$CO_2 + 2H_2 \rightarrow CH_2O + H_2O \quad [6]$$

Biogenic methane can migrate out of the zone of methanogenesis and become an organic carbon source (electron donor) for heterotrophic bacteria elsewhere. This redox zonation going from aerobic → Fe–Mn–oxide reduction → sulfate reduction → methanogenesis is a universal phenomenon on the present-day Earth, as demonstrated, for example, in marine sediment pore waters (e.g., Drever, 1997).

Since prokaryotes can only transport soluble compounds across their cell envelopes, they have evolved various metabolic ways in which to obtain nutrients from solid substrates, for example, Fe or Mn oxides, to support their growth. The ability to alter the chemistry of their surrounding environment on the scale of 10s to 100s of nanometers (Purcell, 1977) is what enables prokaryotes to promote the dissolution of otherwise 'stable' mineral phases (based on bulk fluid chemistry) or the precipitation of minerals under apparently undersaturated conditions. Remarkably, it has been recently demonstrated that prokaryotes can produce electrically conductive nanowires under electron acceptor-limiting conditions (Gorby et al., 2006), thus demonstrating that biogeochemical cycling, which is not directly evident from fluid chemistry or mineralogy, may still be possible via interspecies electron transfer (Gorby et al., 2006).

10.14.2.4 The Growth of Microorganisms

Their small size (submicrometer to micrometer-scale) and resulting high surface area to volume ratio of bacteria help facilitate their growth. Since access to nutrients is based on diffusion and active uptake transporters at the cell surface, the greater the surface to volume ratio, the greater the capacity to obtain nutrients (Koch, 1996; Pirie, 1973; Purcell, 1977). Also, since individual prokaryotes only possess a mass of $\sim 1 \times 10^{-12}$ g (wet weight; Neidhardt et al., 1990), extremely low amounts of dissolved nutrients can support their growth and survival. On Earth, prokaryotes are critical components of the food chain because they utilize dilute concentrations of inorganic and organic nutrients that are unavailable directly to higher eukaryotic organisms (Fenchel and Jorgensen, 1977). The ability of prokaryotes to metabolize organic compounds in natural systems is dependent on the chemical properties and the composition and mineralogy of the soil (Knaebel et al., 1994). For example, Scow and Hutson (1992) found significant decreases in microbial degradation of organic carbon in the presence of kaolinite, owing to the effects of sorption and reduced diffusion from the clay surface.

While prokaryotes can be planktonic, attachment to a mineral surface is advantageous in that it allows them to persist or hold their position within a presumably, favorable geochemical flow, letting the nutrients come to them (**Figure 5**; Emerson and Revsbech, 1994; Purcell, 1977). Individual prokaryote species often grow as microcolonies (Hall-Stoodley et al., 2004), tens of microns in diameter, and can occur in association with other microcolonies within biofilms. Where freshwater–-hypersaline aquatic conditions are stable for a long periods of time and, where nutrients and favorable redox conditions exist for rapid growth, these 'biofilms' can reach large dimensions, up to centimeters or even meters in plan view and a mm or more in thickness, forming significant microbial mats on sediment/rock surfaces (**Figure 4**; Nealson and Stahl, 1997).

These mats are characterized by the microbial production of copious quantities of extracellular polymeric substances (EPS or capsule, **Figure 1**; Moser et al., 2003; Wanger et al., 2006). EPS serve many purposes, including protecting the colonies from the 'external' environment and providing a unique microenvironment or habitable zone in which the organisms can control the redox and pH conditions (Roden et al., 2004).

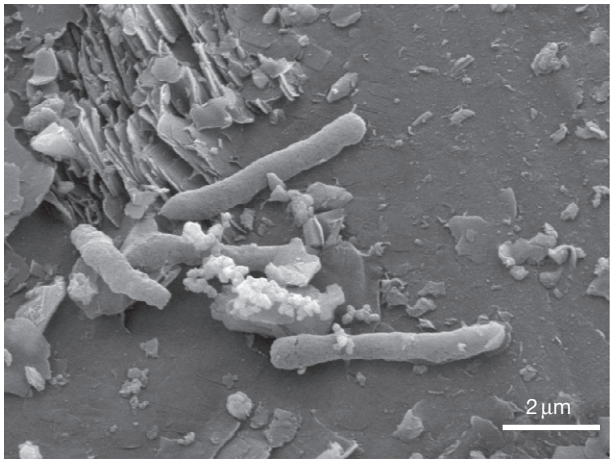


Figure 5 A scanning electron micrograph of a natural fracture face collected 2.8 km below land surface demonstrated that the deep subsurface possesses very sparse bacterial 'biofilms' (see Wanger et al., 2006).

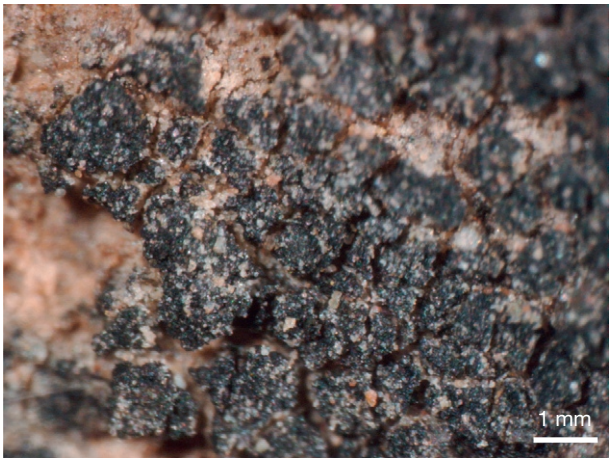


Figure 7 A photograph of an endolithic bacterial colonization (the blackening) of aeolian sandstone in potholes at Bartlett Wash, near Moab, NM (see Chan et al., 2005).

Figure 6 A photograph of microbial mats growing near Atlin, BC, Canada. Prior to the evolutionary development of predators, approximately one-half billion years ago, much of the photic biosphere on Earth presumably looked similar to this.

EPS-rich biofilms can also play an important role in the stabilization of sediment surfaces by cementing mineral particles together (see **Figures 1** and **3**; Stahl, 1994), thus prolonging the stability of a particular habitat.

Under arid environmental conditions where water activity is limited, for example, hot and cold desert environments (Friedmann and Ocampo, 1976), habitable endolithic microenvironments are found within soil or within rock. Endolithic microbial communities display an intimate relationship with the substrate they inhabit, also producing protective and nutritive EPS (**Figure 7**).

10.14.3 Geology and Life

10.14.3.1 Lithotrophy

Lithotrophic or 'rock-eating' prokaryotes are broadly characterized as organisms capable of utilizing inorganic compounds, for example, H_2S or minerals to generate energy, and can produce geologically significant phenomena. For instance, metal sulfide oxidation produces acidic environments as in acid mine drainage systems (Nordstrom and Southam, 1997; Singer and Stumm, 1970) and supergene copper deposits (Enders et al., 2006).

Microorganisms are also active agents of chemical weathering of silicate minerals (e.g., Brantley et al., 2001; Hiebert and Bennett, 1992; Rogers et al., 1998; **Figure 3**), and if reactive organic matter is abundant, mineral dissolution can occur very rapidly. Silica weathering (both as grains and as cement) is a general phenomenon resulting from microbial corrosion in order to concentrate elements limited in their environment, for example, iron and phosphate (Banerjee et al., 2000), and micronutrients, such as Fe, Co, Zn, Mo, Cu, and Ni, which are represented in trace amounts in most terrestrial materials (Bennett et al., 1996). Moreover, if apatite occurs in a silicate rock, it will be the first mineral to be dissolved out because phosphate is a biolimiting nutrient (Taunton et al., 2000). An added advantage of silicate weathering is the creation of suitable habitats for the growth of endolithic microbial communities in water-limited ecosystems (Chan et al., 2005; Friedmann, 1980, 1982). This process also results in the exfoliation of silicate rocks that is typical of hot or cold desert environments, whereby the metabolic activity of microbes inhabiting the protective microhabitat beneath the exposed surface of the rocks weakens the rock in layers parallel to its surface (Parnell et al., 2004).

The importance of mineral weathering in supporting the biosphere has led to the general hypothesis that life on Earth or any other rocky planetary body would ultimately be limited in the long term without some form of plate tectonics or lithospheric renewal providing nutrients from the more reduced lithosphere below. However, once established, subsurface colonization of planetary bodies would extend life for much longer periods of time as the planet cools, allowing the deeper penetration of life into the lithosphere. For example, Hoening et al. (2013) modeled the evolution of the continental surface coverage and the mantle water content with and without

biologically enhanced weathering rates. These authors speculated that the mantle would be significantly drier and the aerial surface coverage by continents smaller than observed if the Earth had no biosphere. The Earth may even lack plate tectonics altogether under these circumstances!

10.14.3.2 Microorganisms and Mineral Formation

Given sufficient mineral ions in the aqueous medium, microorganisms and their biofilms or mats can be rapidly fossilized (**Figure 2**; Schultze-Lam et al., 1995; Westall et al., 1995). Fossilization takes place either (i) by the complexation of mineral ions to functional groups in the organic material, gradually replacing the organic structure and, at the same time, trapping the degrading organic molecules in the polymerizing mineral matrix, or (ii) when amorphous minerals formed in the aqueous phase bind to the cell envelope (Konhauser et al., 2003; Rancourt et al., 2005). The organic matter in the fossilized microorganisms becomes rapidly degraded (e.g., Orange et al., 2013), but the degradation products may remain trapped in the encapsulating mineral matrix. Alternatively, in modern oxidizing hydrothermal environments, the organic matter can be completely oxidized, although the microbial cast still remains (Konhauser, 1998).

The formation of many secondary minerals catalyzed by microorganisms occurs in both natural and laboratory systems (**Figure 8**; see Lowenstam, 1981). Metal sorption and mineral nucleation onto microbial cell envelopes are two different passive mineralization processes. Metal sorption occurs via ion-exchange reactions through competition between hydronium ions, alkaline earth ions, and heavy metals for anionic reactive sites on microbial surfaces (Beveridge, 1981; Urrutia and Beveridge, 1994). Many cell envelope polymers have a demonstrated capacity to bind metallic ions and promote mineral nucleation (Beveridge, 1981; Fein et al., 1997).

Active surface-mediated mineralization occurs either by the direct transformation of metals into unstable forms (Lovley et al., 1991) or by the formation of metal-reactive by-products (Fortin et al., 1994). For example, cyanobacteria have the capacity to precipitate carbonates (Thompson and Ferris, 1990; Thompson et al., 1997) via the creation of alkaline conditions immediately around the cell from photosynthetic activity (**Figure 6**). However, the formation of unstable forms or metal-reactive by-products depends on the geochemical environment in which they are formed. One hypothesis for the deposition of banded iron formations in the late Archean is that iron oxide precipitation was mediated by the metabolic activity of oxygenic photosynthetic cyanobacteria (although there are a number of BIFs that formed by abiotic processes on the early, anoxic Earth before the advent of oxygenic photosynthesis). Today and since the oxygenation of Earth's atmosphere, the oxidation of iron (reaction [7]) occurs within the aerobic/anaerobic interface of the Earth surface where the reduced iron encounters oxidizing conditions:

$$Fe^{2+} + \tfrac{1}{4}O_2 + H^+ \rightarrow Fe^{3+} + \tfrac{1}{2}H_2O,\ \Delta G^o = -40.1\,kJ\,mol^{-1} \quad [7]$$

(Edwards, et al., 2004; Ehrlich, 1975; Emerson and Revsbech, 1994; Little et al., 2004). Dissimilatory metal-reducing prokaryotes are best known for their ability to utilize minerals for their electron acceptors resulting in the solubilization of Fe (Lovley and Phillips, 1988) and Mn (Myers and Nealson, 1988) oxides and any coprecipitated base metals, as well as the formation of magnetite, siderite, chromium hydroxide, and uraninite (Ishibashi et al., 1990; Lovley et al., 1991; Pedersen and Ekendahl, 1990). The most common metal sulfide attributed to biogenic activity is iron sulfide (Tuttle et al., 1969; reaction [8]):

$$Fe^{2+} + H_2S \rightarrow FeS(s) + 2H^+ \quad [8]$$

These dissimilatory processes combined with bacterial surface catalysis are responsible for mineral formation on cell surfaces in these systems.

Although individual bacteria are extremely small, bacteria have played a role in the formation of geologically significant deposits of carbonate, banded iron formations, and phosphorites (Konhauser et al., 2002; Riding, 2000; Youssef, 1965; see **Figure 9**). From a geochemical perspective, the contribution of bacteria to the formation of ancient deposits is often inferred by the enrichment of stable light isotopes by microbial mineral-forming processes (Kakegawa, 2001; Schidlowski, 1988). Mineral signatures then can serve as biomarkers when they are associated with fossils (Cady et al., 2003; Rasmussen, 2000; Southam et al., 2001; Tazaki et al., 1992), when they possess isotopic signatures that point to biochemical (enzymatic) processing of dissolved solutes, or when realistic, abiotic geologic phenomena cannot explain their occurrence (see review in Westall and Southam, 2006).

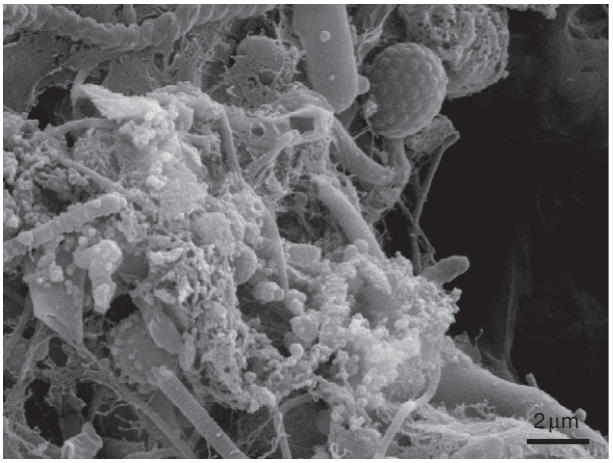


Figure 8 A scanning electron micrograph of a biofilm collected from a borehole flowing from the deep subsurface, 3.1 km below land surface in the Witwatersrand Basin, Republic of South Africa. Note the fine-grained, amorphous iron silicate minerals that have precipitated within the biofilm.

10.14.3.3 Habitable Conditions on Early Earth (and Other Rocky Planets)

As we saw briefly in the Introduction, the early Earth would have presented a very different habitat compared with the modern planet (see reviews in Nisbet and Fowler, 2004; Nisbet and Sleep, 2001). The hotter mantle meant that there was a greater amount of volcanic and hydrothermal activity (perhaps similar

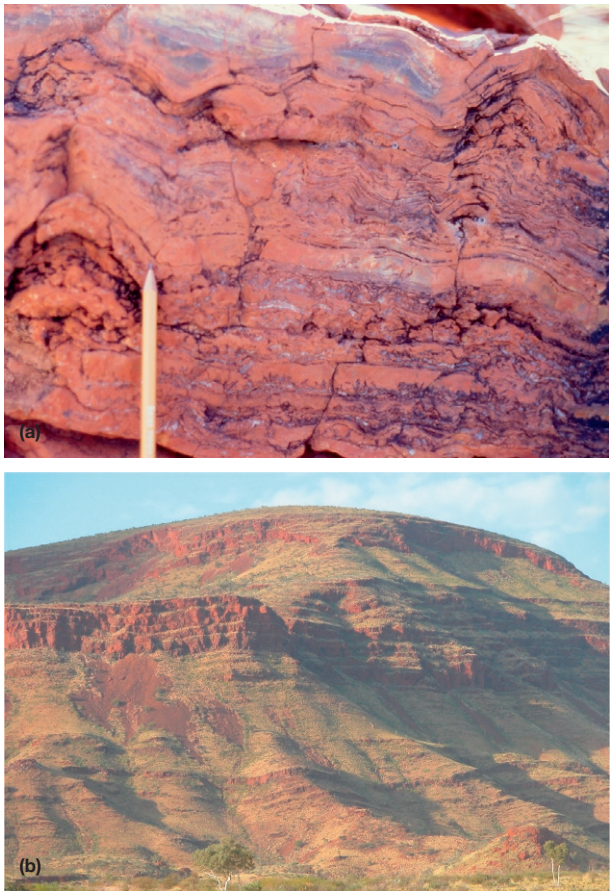


Figure 9 (a) Small, conical stromatolites from the early Archean (>3.46 billion years old) North Pole Dome of the Pilbara (photograph D. Marchesini). (b) A photograph of the Brockman Formation, a banded iron formation found in the Hamersley Range, Pilbara, Western Australia.

to that observed in the central portion of **Figure 2**), as well as faster recycling of the crust (Arndt, 1994). The atmosphere was composed of mostly CO_2, together with CO, NH_3, N_2, and probably a certain amount of CH_4 (the latter possibly at least partly biological; Pavlov et al., 2001; Zahnle et al., 2010). Oxygen levels were low (<0.2% of present atmospheric levels, Kasting, 1993), the oxygen being produced by photolysis of water vapor in the upper atmosphere. Based on the high CO_2 concentrations in the atmosphere (Rye et al., 1995), the oceans were probably slightly acidic, pH 6–7 (Grotzinger and Kasting, 1993). These acidic conditions, combined with the rapid cycling of hydrothermal fluids (Duchač and Hanor, 1987; Hofmann and Bolhar, 2007; Paris et al., 1985), would have resulted in extensive mineral weathering. Higher salinity (de Ronde et al., 1997) and somewhat higher temperatures (≤40 °C according to Hren et al., 2009) were also characteristic of the early to mid-Archean oceans. However, isotopic studies of fluid inclusions in Archean rocks have found that the composition of the oceans has remained somewhat constant over time (Holland, 1984; Kasting et al., 1993). The lack of significant oxygen in the atmosphere meant that UV radiation levels were high, that is, up to 50 times present levels and up to 1000 times in a worst-case scenario (Cockell and Raven, 2004). However, the deleterious effects of UV radiation on the biosphere may have been mitigated by large amounts of dust and aerosols in the atmosphere from volcanic eruptions, meteoritic impacts (Lowe et al., 2003), and possibly water vapor (Lammer et al., 2005) or CH_4 smog (Lovelock, 1988).

The coexistence of reducing and oxidized gases in the prebiotic atmosphere, combined with the input of extraterrestrial organic carbon (e.g., Brown et al., 2000), would have provided abundant untapped chemical energy for early life-forms (Catling and Claire, 2005), for example, methanogens that utilize the redox disequilibrium by combining H_2 and CO_2 to make methane (reaction [4]).

10.14.3.4 The Earliest Traces of Life

We would describe the early life-forms on Earth as being anaerobic, halotolerant, UV-resistant (see Daly et al., 1994), and thermophilic (see Stetter et al., 1990). What is intriguing about the remnants of the oldest life on Earth is (1) the modern-looking aspect of the individual microorganisms, their colonies, and growth as biofilms and mats (**Figure 10**), (2) their diversity, and (3) their wide environmental distribution (see review in Westall and Southam, 2006; Westall, 2011). While stromatolitic structures and shallow water microbial mats of presumably anaerobic photosynthetic origin have been described from the early Archean period (Allwood et al., 2006; Hofmann et al., 1999; Tice and Lowe, 2004; Walsh, 1992; Westall et al., 2006b, 2011), they are orders of magnitude smaller (**Figure 9(a)**) than the aerobic bioconstructions of the late Archean/Proterozoic epochs (**Figure 9(b)**; Awramik and Sprinkle, 1999). Other carbon-rich sediments from this period testify to a relatively abundant input of carbonaceous material from shallower water environments (Tice and Lowe, 2004; Walsh and Lowe, 1999) or from the planktonic photic zone (Javaux et al., 2010; Walsh and Lowe, 1999). However, the main limitation to early Archean productivity appears to be the fact that oxygenic photosynthesis and the corresponding aerobic heterotrophs had not yet developed (Des Marais, 2000).

The advanced level of evolution demonstrated by the remnants of life at 3.5 Ga has implications for the timing of the origin of life. Did the predicted planet-sterilizing impacts (Nisbet and Sleep, 2001; Ryder, 2002; Sleep et al., 1989) during the LHB between 4.0 and 3.85 Ga (Maher and Stevenson, 1988) completely wipe out all traces of life, if it had evolved before, or was life able to survive in protected subsurface ecosystems? Note that an alternative hypothesis proposes that only the topmost 400 m of the oceans would have been evaporated during the worst of the impact events (Ryder, 2002). The occurrence of hyperthermophilic microorganisms in the lowest branches of the evolutionary tree of life (Pace, 1997) is generally considered to imply that early life must have passed through severe (catastrophic) environmental conditions, such as an almost planet-sterilizing impact and that life-forms in protected habitats, such as deep-sea hydrothermal vents, survived, whereas life-forms in the more exposed habitats did not. Whatever the timing necessary for life to appear (Lazcano and Miller, 1996, suggested 10 My), a rocky planet with a hydrosphere can support prokaryote life-forms.

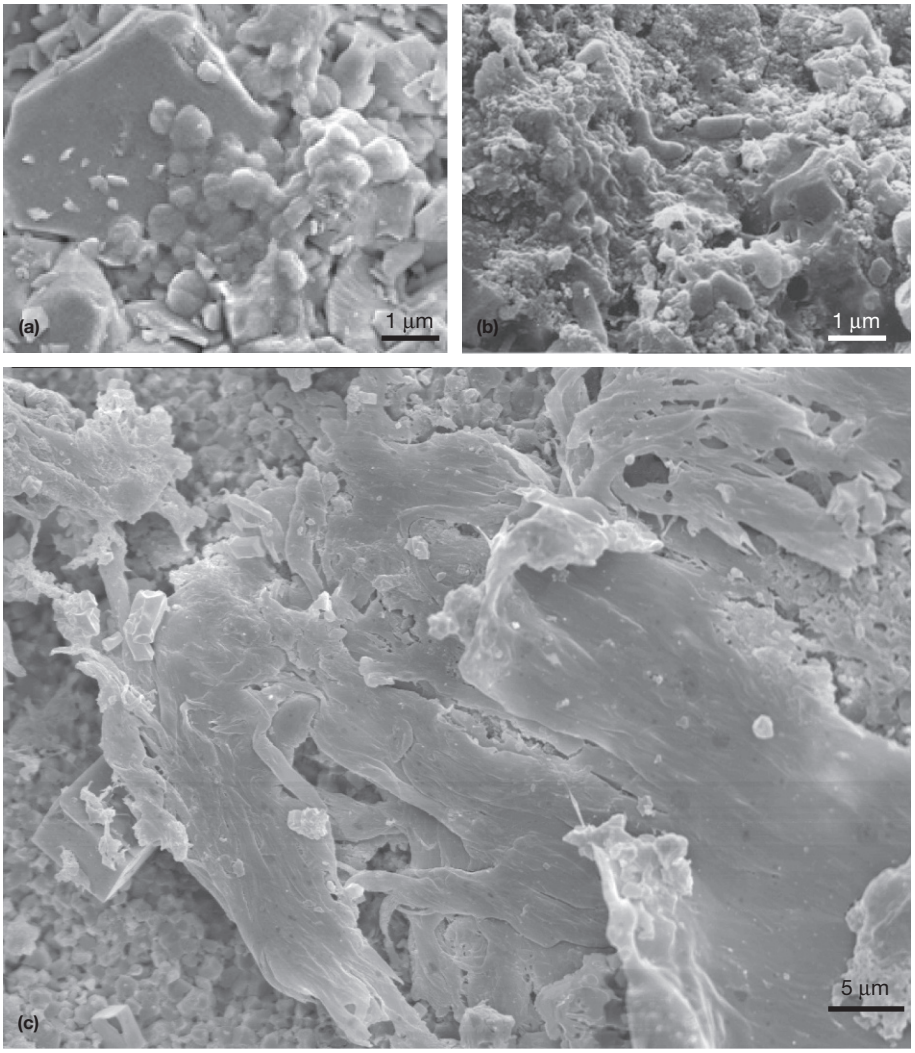


Figure 10 Fossilized early Archean microorganisms (3.47–3.34 billion years) (a) small colony of small coccoid microfossils (arrow) that colonized a marine, subsurface sedimentary habitat (see Westall et al., 2006a); (b) small colony of rod-shaped to vibrioid-shaped microfossils (arrow) that were closely associated with anaerobic, photosynthetic filaments forming a well-developed microbial mat (c) on beach sediments in an evaporitic environment (see Westall et al., 2006b).

10.14.4 Conclusions for Geology, Life, and Habitability Beyond the Earth

The ingredients necessary for life to originate (liquid water in contact with rocks/minerals; abundant organic molecules (exogenous or endogenous); elements such as hydrogen, oxygen, nitrogen, phosphorous, and sulfur; and energy sources (geochemical and sunlight)) were abundant on the early Earth. They were also abundant elsewhere in the early solar system and presumably in the universe. It is therefore legitimate to surmise that life could have arisen independently wherever these initial conditions are stable for the length of time that is necessary for the process. Once started, from a microbiological point of view, these 'primitive' life-forms can exist wherever there are the essential nutrients and potential redox reactions (providing chemical energy) that could support growth. Apart from the classical terrestrial planets in the solar system (Venus and Mars), Europa has also recently excited interest because of the potential for liquid water in contact with a rocky core (Bhattacherjee and Chela-Flores, 2004; Greenberg, 2005), and the presence of water ice on Enceladus is equally stimulating (Brown et al., 2006).

The possibility of extraterrestrial life on Mars has fascinated humans for over a hundred years: Percival Lowell (1910) started the ball rolling with an account of Martian life based on Schiaparelli's (1879) drawing of what he thought were channels (or intersecting straight lines) on the surface of Mars. Although various missions over the last 30 years have discounted these fanciful theories, it is clear that, early in its history, there was a significant inventory of water on the surface of the planet as documented by geomorphological features (e.g., Di Achille and Hynek, 2010; Jaumann et al., 2002) and the presence of minerals produced by aqueous alteration and deposition (e.g., Bibring et al., 2006; Loizeau et al., 2012), albeit probably not sufficient for an ocean covering the northern hemisphere (Clifford and Parker, 2001;

Mouginot et al., 2012), that is, Mars possessed an active hydrosphere during the time when life originated on Earth. As we have seen earlier, microbial life does not require oceans of water in order to function: Once it has got going, it just requires continual access to a very small amount of carbon, nutrients, and energy. Given the access to the basic requirements for an origin of life (liquid water, organics, and rocks/minerals), it is possible that life did arise independently on the red planet (Westall et al., 2013). Indeed, the first results from the Mars Science Laboratory mission with its rover Curiosity in the 3.6-Ga-old Gale crater indicate that there is evidence for habitable conditions in the crater (http://www.nasa.gov/mission_pages/msl/news/msl20130312.html). Therefore, Mars may contain a record of fossilized primitive microorganisms or their associated mineralogical biomarkers, similar to those that are found on Earth. These traces of early Martian life will likely occur in the rocks of the southern hemisphere, which are largely Noachian in age and have apparently undergone aqueous alteration. Of course, boulders or meteorites of southern hemisphere material could be found elsewhere on Mars or even on the moons of Mars, Phobos, and Demos, since Mars, together with the other inner solar system planets, experienced a period of late bombardment between 4.0 and 3.8 billion years ago. It is equally possible that the Noachian-age rocks will contain traces of the prebiotic stage of life and the appearance of the first cells, that is, the critical period of life on Earth's history that is missing on Earth – if we are capable of recognizing these traces (Westall et al., 2013).

If life established itself on Mars during the Noachian age, could it still be present today and, if so, where? The loss of liquid water on Mars' surface and the cessation of any sort of plate tectonic activity (if there ever was any) due to the rapid cooling of the planet limit one of the important criteria for life, that is, a flux of reduced or energetic redox species, although Curiosity did find evidence of the redox couple sulfide and sulfate in Gale crater (Grotzinger et al., 2014; McLennan et al., 2014). As we noted earlier, microorganisms can survive a wide range of extreme conditions, but there are certain combinations that are lethal, and it appears that the combined low water activity, high UV radiation, and the presence of some kind of oxidant species that oxidizes labile organic components (i.e., those that are necessary for life) makes the surface of the planet hostile to any long-term survival. However, Mars is not completely volcanically inactive (Neukum et al., 2004), and there could be hydrothermal activity in its subsurface, as indicated by occasional surface exposure (Malin and Edgett, 2000) and with subsequent implications for life (Farmer, 1996; Schwenzer et al., 2012). Previous identifications of methane on Mars (e.g., Formisano et al, 2004) gave rise to hypotheses of a subsurface origin related to biological activity (Tung et al., 2005) although an alternative hypothesis related hydrothermal melting of CH_4 clathrates was also postulated (Prieto-Ballesteros et al., 2006). However, it now appears that the detections of methane were erroneous (Zahnle et al., 2011). Whether or not methane has been detected, this activity would provide the necessary criteria/conditions for life, at least in the subsurface of Mars, similar to that provided to bacteria in Earth's subsurface (Boston et al., 1992; Wanger et al., 2006).

While we do not know the exact conditions necessary to create life, the basic requirements of a habitable extraterrestrial body (assuming Earth is a representative sample) are fairly well understood. Habitable locations in our solar system today, that is, places where life can be sustained, are not as restrictive as those where life could originate and include Mars' subsurface (Boston et al., 1992), as well as the ocean on Europa (Chyba and Phillips, 2001). It is also intriguing to realize that, in the future as the Sun gets hotter and comes into the red giant phase of its history, these planets and possibly others in the outer solar system will be able to host life, if not see it appear independently. Thus, the possibility of finding traces of extinct or extant life on other planets in the solar system is strong, and the future for life in the solar system is optimistic, even though the habitability of our own planet will slowly decrease starting in another c.1.5 Ga (Franck et al., 2006).

References

Allwood AC, Walter MR, Kamber BS, Marshall CP, and Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* 441: 714–718.

Anderson IK, Ashton JH, Boyce AJ, Fallick AE, and Russell MJ (1998) Ore depositional processes in the Navan Zn-Pb deposit Ireland. *Economic Geology* 93: 535–563.

Arndt NT (1994) Archean komatiites. In: Condie KC (ed.) *Archean Crustal Evolution*, pp. 11–44. Amsterdam: Elsevier.

Atlas RM and Bartha R (1997) *Microbial Ecology: Fundamentals and Applications*, 4th edn. Menlo Park, CA: Benjamin Cummings, 306 p.

Awramik SM and Sprinkle J (1999) Proterozoic stromatolites: The first marine evolutionary biota. *Historical Biology* 13: 241–253.

Banerjee M, Whitton BA, and Wynn-Williams DD (2000) Phosphatase activities of endolithic communities in rocks of the Antarctic dry valleys. *Microbial Ecology* 39: 80–91.

Barbier B and Brack A (2002) Conformation-controlled hydrolysis of polyribonucleotides by sequential basic polypeptides. *Journal of the American Chemical Society* 114: 3511–3515.

Barns SM and Nierzwicki-Bauer SA (1997) Microbial diversity in ocean, surface and subsurface environments. In: Banfied JF and Nealson KH (eds.). Geomicrobiology: Interactions Between Microbes and Minerals. *Reviews in Mineralogy* 35: 35–79.

Bechtel A, Sieh Y-N, Pervaz M, and Puttmann W (1996) Biodegradation of hydrocarbons and biogeochemical sulfur cycling in the salt dome environment: Inferences from sulfur isotope and organic geochemical investigations of the Bahloul formation at the Bou Grine Zn/Pb ore deposit (Tunisia). *Geochimica et Cosmochimica Acta* 60: 2833–2855.

Bennett PC, Hiebert FK, and Choi WJ (1996) Rates of microbial weathering of silicates in ground water. *Chemical Geology* 132: 45–53.

Beveridge TJ (1981) Ultrastructure, chemistry, and function of the bacterial wall. *International Review of Cytology* 72: 229–317.

Bhattacherjee AB and Chela-Flores J (2004) Search for bacterial waste as a possible signature of life on Europa. In: Seckbach J, Chela-Flores J, Owen T, and Raulin F (eds.). *Life in the Universe, Cellular Origin and Life in Extreme Habitats and Astrobiology*, vol. 7, pp. 257–260. Dordrecht, The Netherlands: Springer.

Bibring J-P, Langevin Y, Mustard J-F, et al. (2006) Global mineralogical and aqueous Mars history derived from OMEGA / Mars Express Data. *Science* 312: 400–403.

Boston PJ, Ivanov MV, and McKay CP (1992) On the possibility of chemosynthetic ecosystems in subsurface habitats on Mars. *Icarus* 95: 300–308.

Brack A (2004) The chemistry of life's origins. In: Seckback J (ed.) *Life in the Universe*, pp. 59–73. Dordrecht: Kluwer.

Brack A (2005) L'exobiologie ou l'origine chimique de la vie. In: Halbwachs JL, Egret D, and Hameury JM (eds.) *Formation planétaire et exoplanètes*, *Ecole CNRS de Goutelas*, XXVIII, pp. 1–24.

Brantley SL, Lierman L, Bau M, and Wu S (2001) Uptake of trace metals and rare elements from hornblende by a soil bacterium. *Geomicrobiology Journal* 18: 37–61.

Brill WJ (1975) Regulation and genetics of bacterial nitrogen fixation. *Annual Review of Microbiology* 29: 109–129.

Brochier C and Philippe H (2002) A non-hyperthermophilic ancestor for bacteria. *Nature* 417: 244.

Brown RH, Clark RN, Buratti BJ, et al. (2006) Composition and physical properties of Enceladus' surface. *Science* 311: 1425–1428.

Brown ME and Hand KP (2013) Salts and radiation products on the surface of Europa. *The Astronomical Journal* 145: 110. http://dx.doi.org/10.1088/0004-6256/145/4/110.

Brown PG, Hildebrand AR, Zolensky ME, et al. (2000) The fall, orbit, and composition of the Tagish Lake Meteorite: A new type of carbonaceous chondrite. *Science* 290: 320–325.

Cady SL, Farmer JD, Grotzinger JP, Schopf WJ, and Steele A (2003) Morphological biosignatures and the search for life on Mars. *Astrobiology* 3: 351–368.

Catling DC and Claire MW (2005) How Earth's atmosphere evolved to an oxic state: A status report. *Earth and Planetary Science Letters* 237: 1–20.

Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, and Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila*: Possible chemoautotrophic symbionts. *Science* 213: 340–342.

Chan M, Moser K, Davis JM, Southam G, Hughes K, and Graham T (2005) Desert potholes: Life in the fast pool. *Aquatic Geochemistry* 11: 279–302.

Chapelle FH (1993) *Ground-Water Microbiology and Geochemistry*. New York: Wiley 448 p.

Chyba CF and Phillips CB (2001) Possible ecosystems and the search for life on Europa. *PNAS* 98: 801–804.

Chyba CF and Sagan C (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origins of life. *Nature* 355: 125–132.

Clifford SM and Parker TJ (2001) The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* 154: 40–79.

Cockell CS (2008) The interplanetary exchange of photosynthesis. *Origins of Life and Evolution of the Biosphere* 38: 87–104.

Cockell CS and Raven JA (2004) Zones of photosynthetic potential on Mars and the early Earth. *Icarus* 169: 300–310.

Colwell FS, Onstott TC, Dilwiche ME, et al. (1997) Microorganisms from deep, high temperature sandstones: Constraints on microbial colonization. *FEMS Microbiology Reviews* 20: 425–435.

Daly MJ, Ouyang L, and Minton KW (1994) *In vitro* damage and *recA*-dependent repair of plasmid and chromosomal DNA in the radioresistant bacterium *Deinococcus radiodurans*. *Journal of Bacteriology* 176: 3508–3517.

de Ronde CEJ, DeR Channer DM, Faure K, Bray CJ, and Spooner ETC (1997) Fluid chemistry of Archean seafloor hydrothermal vents; implications for the composition of circa 3.2 Ga seawater. *Geochimica et Cosmochimica Acta* 61: 4025–4042.

De Vera JPP, Schulze-Makuch D, Khan A, et al. (2013) Adaptation of an Antarctic lichen to Martian niche conditions can occur within 34 days. *Planetary and Space Science* 98: 169–181, http://dx.doi.org/j.pss.2013.07.014.

Deamer DW (1998) Membrane compartments in prebiotic evolution. In: Brack A (ed.) *The Molecular Origins of Life: Assembling Pieces of the Puzzle*, pp. 189–205. Cambridge: Cambridge Univ. Press.

Des Marais DJ (2000) When did photosynthesis emerge on Earth? *Science* 289: 1703–1705.

Di Achille G and Hynek BM (2010) Ancient ocean on Mars supported by global distribution of deltas and valleys. *Nature Geoscience* 3: 459–463. http://dx.doi.org/10.1038/ngeo891.

Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, and Tyryshkin A (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *PNAS* 98: 2170–2175.

Donald R and Southam G (1999) Low temperature anaerobic bacterial diagenesis of ferrous monosulfide to pyrite. *Geochimica et Cosmochimica Acta* 63: 2019–2023.

Drever JI (1997) *The Geochemistry of Natural Waters*, 3rd edn. Upper Saddle River, NJ: Prentice Hall, 436 p.

Duchač KC and Hanor JS (1987) Origin and timing of the metasomatic silicification of an Early Archaean komatiite sequence, Barberton Mountain Land, South Africa. *Precambrian Research* 37: 125–146.

Edwards KJ, Bach W, McCollom TM, and Rogers DR (2004) Neutrophilic iron-oxidizing bacteria in the ocean: Their habitats, diversity, and roles in mineral deposition, rock alteration, and biomass production in the deep-sea. *Geomicrobiology Journal* 21: 393–404.

Ehrlich HL (1975) The formation of ores in the sedimentary environment of the deep sea with microbial participation: The case for ferromanganese concretions. *Science* 119: 36–41.

Emerson D and Revsbech NP (1994) Investigation of an iron-oxidizing microbial mat community located near Aarhus, Denmark: Field studies. *Applied and Environmental Microbiology* 60: 4022–4031.

Enders MS, Southam G, Knickerbocker C, and Titley SR (2006) The role of microorganisms in the supergene environment of the Morenci porphyry copper deposit, Greenlee County, Arizona. *Economic Geology* 101: 59–70.

Farmer J (1996) Hydrothermal systems on Mars: An assessment of present evidence. In: Bock GR and Goode JA (eds.) *Evolution of Hydrothermal Systems on Earth (and Mars?)*, Ciba Foundation Symposium 202, pp. 273–299. Chichester: Wiley.

Fein JB, Daughney CJ, Yee N, and Davis T (1997) A chemical equilibrium model of metal adsorption onto bacterial surfaces. *Geochimica et Cosmochimica Acta* 61: 3319–3328.

Fenchel TM and Jorgensen BB (1977) Detritus food chains of aquatic ecosystems: The role of bacteria. *Advances in Microbial Ecology* 1: 1–58.

Ferris JP, Hill AR, Lui R, and Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 381: 59–61.

Formisano V, Atreya S, Encrenaz T, Ignatiev N, and Giuranna M (2004) Detection of methane in the atmosphere of Mars. *Science* 306: 1758–1761.

Fortin D, Southam G, and Beveridge TJ (1994) An examination of iron sulfide, iron-nickel sulfide and nickel sulfide precipitation by a *Desulfotomaculum* species and its nickel resistance mechanisms. *FEMS Microbiology Ecology* 14: 121–132.

Fox SW and Dose K (eds.) (1977) *Molecular Evolution and the Origin of Life*. New York: Marcel Dekker.

Franck S, Bounama C, and von Bloh W (2006) Causes and timing of future biosphere extinctions. *Biogeosciences* 3: 85–92.

Friedmann EI (1980) Endolithic microbial life in hot and cold deserts. *Origins of Life* 10: 223–235.

Friedmann EI (1982) Endolithic microorganisms in the Antarctic cold desert. *Science* 215: 1045–1052.

Friedmann EI and Ocampo R (1976) Endolithic blue-green algae in the dry valleys: Primary producers in the Antarctic desert ecosystems. *Science* 193: 1247–1249.

Gilichinsky DA, Wilson GS, Friedmann EI, et al. (2007) Microbial populations in Antarctic permafrost: Biodiversity, state, age and implication for astrobiology. *Astrobiology* 7: 275–311.

Goldman M and Wilson DA (1977) Growth of *Sporosarcina ureae* in defined media. *FEMS Microbiology Letters* 2: 113–115.

Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.

Goodwin S and Zeikus JG (1987) Physiological adaptations of anaerobic bacteria to low pH: Metabolic control of proton motive force in *Sarcina ventriculi*. *Journal of Bacteriology* 169: 2150–2157.

Gorby YA, Yanina S, McLean JS, et al. (2006) Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. *PNAS* 103: 11358–11363.

Greenberg R (2005) *Europa – The Ocean Moon: Search for an Alien Biosphere*. Chichester, UK: Springer Praxis Publishing, 380 p.

Grotzinger JP and Kasting JF (1993) New constraints on Precambrian ocean composition. *Journal of Geology* 101: 235–243.

Grotzinger JP, Sumner DY, Kah LC, et al. (2014) MSL Science Team (2014) A habitable fluvio-lacustrine environment at Yellowknife Bay, Gale Crater, Mars. *Science* 343(6169): 386–387. http://dx.doi.org/10.1126/science.1242777.

Hall-Stoodley L, Costerton JW, and Stoodley P (2004) Bacterial biofilms: From the environment to infectious disease. *Nature Reviews. Microbiology* 2: 95–108.

Heising S and Schink B (1998) Phototrophic oxidation of ferrous iron by a *Rhodomicrobium vannielii* strain. *Microbiology* 144: 2263–2269.

Hiebert F and Bennett PC (1992) Microbial control of silicate weathering in organic-rich ground water. *Science* 258: 278–281.

Hoening D, Hansen-Goos H, Airo A, and Spohn T (2013) Biotic vs. abiotic Earth: A model for mantle hydration and continental growth. *Planetary and Space Science* 98: 5–13. http://dx.doi.org/10.1016/j.pss.2013.10.004.

Hofmann A and Bolhar R (2007) The origin of carbonaceous cherts in the Barberton greenstone belt and their significance for the study of early life in mid-Archaean rocks. *Astrobiology* 7: 355–388.

Hofmann HJ, Grey K, Hickman AH, and Thorpe RI (1999) Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Geological Society of America Bulletin* 111: 1256–1262.

Holland HD (1984) *The Chemical Evolution of the Atmosphere and Oceans*. Princeton: Princeton Univ Press.

Holm NG and Charlou J-L (2001) Initial indications of abiotic formation of in the Rainbow ultramafic hydrothermal system, Mid-Atlantic ridge. *Earth and Planetary Science Letters* 191: 1–8.

Hren MT, Tice MM, and Chamberlain CP (2009) Oxygen and hydrogen isotope evidence for a temperate climate 3.42 billion years ago. *Nature* 462: 205–208.

Inoue T and Orgel LE (1982) Oligomerization if (guanosine 5'-phosphor)-2-methylimidazolide on poly(C). *An RNA polymerase model*. *Journal of Molecular Biology* 162: 201–217.

Ishibashi Y, Cervantes C, and Silver S (1990) Chromium reduction in *Pseudomonas putida*. *Applied and Environmental Microbiology* 56: 2268–2270.

Jaumann R, Hauber E, Lanz J, Hoffmann H, and Neukum G (2002) Geomorphological record of water-related erosion on Mars. In: Horneck G and Baumstark-Kahn C (eds.) *Astrobiology*, pp. 89–109. Berlin Heidelberg New York: Springer.
Javaux EJ, Marshall CP, and Bekker A (2010) Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits. *Nature* 463: 934–938.
Johnson AP, Cleaves HJ, Dworkin JP, Glavin DP, Lazcano A, and Bada JL (2008) The Miller volcanic spark discharge experiment. *Science* 322: 404.
Kakegawa T (2001) Isotopic signatures of early life in the Archean oceans: Influence from submarine hydrothermal activities. In: Nakashsima S, Maruyama S, Brack A, and Windley BF (eds.) *Geochemistry and the Origin of Life*, pp. 237–249. Tokyo: Universal Acad. Press.
Kappler A and Newman DK (2004) Formation of Fe(III)-minerals by Fe(II)-oxidizing photoautotrophic bacteria. *Geochimica et Cosmochimica Acta* 68: 1217–1226.
Kashefi K and Lovley DR (2003) Extending the upper temperature limit for life. *Science* 301: 934.
Kasting JF (1993) Earth's early atmosphere. *Science* 259: 920–926.
Kasting JF, Whitmire DP, and Reynolds RT (1993) Habitable zones around main sequences stars. *Icarus* 101: 108–128.
Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.
Knaebel DB, Federle TW, McAvoy DC, and Vestal JR (1994) Effect of mineral and organic soil constituents on microbial mineralization of organic compounds in a natural soil. *Applied and Environmental Microbiology* 60: 4500–4508.
Knauth LP and Lowe DR (2003) High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geological Society of America Bulletin* 115: 566–580.
Koch AL (1996) What size should a bacterium be? A question of scale. *The Annual Review of Microbiology* 50: 317–348.
Konhauser K (1998) Diversity of bacterial iron mineralization. *Earth Science Reviews* 43: 91–121.
Konhauser KO, Hamade T, Morris RC, et al. (2002) Could bacteria have formed the Precambrian banded iron formations? *Geology* 30: 1079–1082.
Konhauser KO, Jones B, Reysenbach A-L, and Renaut RW (2003) Hot spring sinters: Keys to understanding Earth's earliest life forms. *Canadian Journal of Earth Sciences* 40: 1713–1724.
Kopparapu RK, Ramirez R, Kasting JF, et al. (2013) Habitable zones around main-sequence stars: New Estimates. *The Astrophysical Journal*, arXiv:1301.6674v2 [astro-ph.EP].
Kunte HJ (2006) Osmoregulation in bacteria: Compatible solute accumulation and osmosensing. *Environmental Chemistry* 3: 94–99.
Lammer H, Yu N, Kulikov T, et al. (2005) Stellar-planetary relations: Atmospheric stability as a prerequisite for planetary habitability. In: Dvorak R and Ferraz-Mello S (eds.) *A Comparison of the Dynamical Evolution of Planetary Systems*, *Proceedings of the 6th Alexander von Humboldt colloquium*, Boston, MA: Kluwer.
Lazcano A and Miller SL (1996) Astrophysical and astrochemical insights into the origin of life. *Cell* 85(6): 793–798.
Little CTS, Glynn SEJ, and Mills RA (2004) Four-hundred-and-ninety-million-year record of bacteriogenic iron oxide precipitation at sea-floor hydrothermal vents. *Geomicrobiology Journal* 21: 415–429.
Ljungdahl LG and Eriksson K-E (1985) Ecology of microbial cellulose degradation. *Advances in Microbial Ecology* 8: 237–299.
Loizeau D, Carter J, Bouley S, et al. (2012) Characterization of hydrated silicate-bearing outcrops in Tyrrhena Terra, Mars. *Icarus* 219: 476–497.
Lovelock JE (1988) *The Ages of Gaia*. New York: W.W. Norton.
Lovley DR and Chapelle FH (1995) Deep subsurface microbial processes. *Reviews of Geophysics* 33: 365–381.
Lovley DR and Klug MJ (1983) Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. *Applied and Environmental Microbiology* 45: 187–192.
Lovley DR and Phillips EJP (1988) Novel mode of microbial energy metabolism: Organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Applied and Environmental Microbiology* 54: 1472–1480.
Lovley DR, Phillips EJP, Gorby YA, and Landa ER (1991) Microbial reduction of uranium. *Nature* 350: 413–416.
Lowe DR and Byerly GR (1986) Early Archean silicate spherules of probable impact origin, South Africa and Western Australia. *Geology* 14: 83–86.
Lowe DR, Byerly GR, Kyte FT, Shukulyukov A, Asaro F, and Krull A (2003) Spherule beds 3.47-3.24 billion years old in the Barberton Greenstone Belt, South Africa: A record of large meteorite impacts and their influence on early crustal and biological evolution. *Astrobiology* 3: 7–48.
Lowell P (1910) *Mars as the Abode of Life*. New York: Macmillan, 288 p.
Lowenstam HA (1981) Minerals formed by organisms. *Nature* 211: 1126–1131.
Maher KA and Stevenson DJ (1988) Impact frustration of the origin of life. *Nature* 331: 612–614.
Malin MC and Edgett KS (2000) Evidence for recent groundwater seepage and surface runoff on Mars. *Science* 288: 2330–2335.
Marshall KC (1988) Adhesion and growth of bacteria at surfaces in oligotrophic habitats. *Canadian Journal of Microbiology* 34: 503–506.
Maurette M, Duprat J, Engrand C, et al. (2000) Accretion of neon, organics, CO_2, nitrogen and water from large interplanetary dust particles on the early Earth. *Planetary and Space Science* 48: 1117–1137.
McLennan SM, Anderson RB, Bell JF III, et al. (2014) MSL Science Team (2014) Elemental geochemistry of sedimentary rocks at Yellowknife Bay, Gale Crater, Mars. *Science* 343(6169): 1244734. http://dx.doi.org/10.1126/science.1244734.
Miller SL (1953) A production of amino acids under possible primitive earth condition. *Science* 117: 528–529.
Moser DP, Onstott TC, Fredrickson JK, et al. (2003) Evolution of microbial community structure and geochemistry in an ultradeep South African gold mine borehole. *Geomicrobiology Journal* 20: 517–548.
Mouginot J, Pommerol A, Beck P, Kofman K, and Clifford SM (2012) Dielectric map of the Martian northern hemisphere and the nature of plain filling materials. *Geophysical Research Letters* 39: L02202. http://dx.doi.org/10.1029/2011GL050286.
Myers CP and Nealson KN (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* 240: 1319–1321.
Mykytczuk NCS, Foote SJ, Omelon CR, Southam G, Greer WC, and Whyte LG (2013) Bacterial growth at −15 °C; Molecular insights from the permafrost bacterium *Planococcus halocryophilus* Or1. *International Society for Microbial Ecology* 7: 1211–1226.
Navarro-Gonzalez R, Rainey FA, Molina P, et al. (2003) Mars-like soils in the Atacama Desert, Chile, and the dry limit of microbial life. *Science* 302: 1018–1021.
Nealson KH and Stahl DA (1997) Microorganisms and biogeochemical cycles: What can we learn from layered microbial communities? In: Banfield JF and Nealson KH (eds.) *Geomicrobiology: Interactions Between Microbes and Minerals*. *Rev. Mineral*, 35, pp. 5–34.
Neidhardt FC, Ingraham JL, and Schaechter M (1990) *Physiology of the Bacterial Cell – A Molecular Approach*. Sunderland, MA: Sinauer Associates Inc, 506 p.
Neukum G, Jaumann R, Hoffmann H, et al. (2004) Mars: Recent and episodic volcanic, hydrothermal, and glacial activity revealed by the Mars Express High Resolution Stereo Camera (HRSC) experiment. *Nature* 432: 971–979.
Newman MJ and Rood RT (1977) Implications of solar evolution for the Earth's early atmosphere. *Science* 198: 1035–1037.
Nisbet EG and Fowler CMR (2004) The early history of life. In: Holland HD, Turekian KK, and Schelsinger WH (eds.) *Treatise on Geochemistry*. *Biogeochemistry*, vol. 8, pp. 1–39. Oxford: Elsevier-Pergamon.
Nisbet EG and Sleep NH (2001) The habitat and nature of early life. *Nature* 409: 1083–1091.
Nordstrom DK and Southam G (1997) Geomicrobiology of sulfide mineral oxidation. *Reviews in Mineralogy* 35: 361–390.
Onstott TC, Lin L-H, Davidson M, et al. (2006) The origin and age of biogeochemical trends in deep fracture water of the Witwatersrand Basin, South Africa. *Geomicrobiology Journal* 23: 369–414.
Orange F, Disnar JR, Gautret P, et al. (2013) Preservation and evolution of organic matter during experimental fossilisation of the hyperthermophilic Archaea *Methanocaldococcus jannaschii*. *Origins of Life and Evolution of Biospheres* 42: 587–609.
Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276: 734–740.
Paris I, Stanistreet IG, and Hughes MJ (1985) Cherts of the Barberton greenstone belt interpreted as products of submarine exhalative activity. *Journal of Geology* 93: 111–129.
Parkes RJ, Cragg BA, Bale SJ, et al. (1994) Deep bacterial biosphere in Pacific Ocean sediments. *Nature* 371: 410–411.
Parnell J, Cockell P, Lee CS, and Osinski GR (2004) Microbial colonization in impact generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulphates on Mars. *International Journal of Astrobiology* 3: 247–256.
Pavlov AA, Kasting JF, Brown LL, Rages KA, and Freedman R (2001) Greenhouse warming by CH_4 in the atmosphere of early earth. *Journal of Geophysical Research* 105: 11981–11990.
Pedersen K and Ekendahl S (1990) Distribution and activity of bacteria in deep granitic groundwaters of southeastern Sweden. *Microbial Ecology* 20: 37–52.
Pirie NW (1973) On being the right size. *The Annual Review of Microbiology* 27: 119–131.

Prieto-Ballesteros O, Kargel JS, Fairen AG, Fernandez-Remolar DC, Dohm JM, and Amils R (2006) Interglacial clathrate destabilization on Mars: Possible contributing source of its atmospheric methane. *Geology* 34: 149–152.

Purcell E (1977) Life at low Reynolds number. *American Journal of Physics* 45: 3–11.

Rancourt DG, Thibault P-J, Mavrocordatos D, and Lamarche G (2005) Hydrous ferric oxide precipitation in the presence of nonmetabolizing bacteria: Constraints on the mechanism of a biotic effect. *Geochimica et Cosmochimica Acta* 69: 553–577.

Rasmussen B (2000) Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature* 405: 676–679.

Reysenbach A-L and Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* 296: 1077–1982.

Riding R (2000) Microbial carbonates: The geological record of calcified bacterial–algal mats and biofilms. *Sedimentology* 47: 179–214.

Roden E, Sobolev D, Glazer B, and Luther G (2004) Potential for microscale bacterial Fe redox cycling at the aerobic-anaerobic interface. *Geomicrobiology Journal* 21: 379–391.

Rogers JR, Bennett PC, and Choi WJ (1998) Feldspars as a source of nutrients for microorganisms. *American Mineralogist* 83: 1532–1540.

Rothschild LJ and Mancinelli RL (2001) Life in extreme environments. *Nature* 409: 1092–1101.

Russell MJ, Hall AJ, and Martin W (2010) Serpentinization and its contribution to the energy for the emergence of life. *Geobiology* 8: 355–371.

Ryder G (2002) Mass influx in the ancient Earth-Moon system and benign implications for the origin of life on Earth. *Journal of Geophysical Research* 107: 5022. http://dx.doi.org/10.1029/2001JE001583.

Rye R, Kuo PH, and Holland HD (1995) Atmospheric carbon dioxide concentrations before 2.2 billion years ago. *Nature* 378: 603–605.

Schiapiarelli GV (1878) La carta di Marte, Memoria Prima, Rome.

Schidlowski M (1988) A 3800 million-year isotopic record of life from carbon in sedimentary rocks. *Nature* 333: 313–318.

Schultze-Lam S, Ferris FG, Konhauser KO, and Wiese RG (1995) *In situ* silicification of an Icelandic hot spring microbial mat: Implications for microfossil formation. *Canadian Journal of Earth Sciences* 32: 2021–2026.

Schwartz AW and Orgel LE (1985) Template-directed synthesis of novel, nucleic acid-like structures. *Science* 228: 585–587.

Schwenzer SP, Abramov O, Allen CC, et al. (2012) Puncturing Mars: How impact craters interact with the Martian cryosphere. *Earth and Planetary Science Letters* 335–336: 9–17.

Scow KM and Hutson J (1992) Effect of diffusion and sorption on the kinetics of biodegradation: Theoretical considerations. *Soil Science Society of America* 56: 119–127.

Sephton MA and Botta O (2008) Extraterrestrial organic matter and the origin of life. *Space Science Reviews* 135: 25–35.

Singer PC and Stumm W (1970) Acidic mine drainage: The rate determining step. *Science* 176: 1121–1123.

Sleep NH, Zahnle KJ, Kasting JF, and Morowitz HJ (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* 342: 139–142.

Southam G, Brock C, Donald R, and Röstad A (2001) Pyrite discs in coal: Evidence for fossilized bacterial colonies. *Geology* 29: 47–50.

Stahl LJ (1994) Microbial mats: Ecophysiological model for intertidal benthic micralgal production. In: Krumbein WE, Paterson DM, and Stal LJ (1994) *Biostabilization of Sediments*. Oldenburg: Bibliotheks und Informationssystem de Carl von Ossietzky Universität Oldenburg, Verlag.

Stetter KO, Fiala G, Huber G, Huber R, and Segerer A (1990) Hyperthermophilic microorganisms. *FEMS Microbiology Reviews* 75: 117–124.

Stevens TO, McKinley JP, and Fredrickson JK (1993) Bacteria associated with deep, alkaline, anaerobic groundwaters in southeast Washington. *Microbial Ecology* 25: 35–50.

Taunton AE, Welch SA, and Banfield JF (2000) Microbial controls on phosphate and lanthanide distributions during granite weathering. *Chemical Geology* 169: 371–382.

Tazaki K, Ferris FG, Wiese RG, and Fyfe WS (1992) Iron and graphite associated with fossil bacteria in chert. *Chemical Geology* 95: 313–325.

Thompson JB and Ferris FG (1990) Cyanobacterial precipitation of gypsum, calcite, and magnesite from natural alkaline lake water. *Geology* 18: 995–998.

Thompson JB, Schultze-Lam S, Beveridge TJ, and Des Marais DJ (1997) Whiting events: Biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. *Limnology and Oceanography* 42: 133–141.

Thomsen TR, Finster K, and Ramsing NB (2001) Biogeochemical and molecular signatures of anaerobic methane oxidation in a marine sediment. *Applied and Environmental Microbiology* 67: 1646–1656.

Tice M and Lowe DR (2004) Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* 431: 549–552.

Tung HC, Bramall NE, and Price PB (2005) Microbial origin of excess methane in glacial ice and implications for life on Mars. *PNAS* 102: 18292–18296.

Tuttle JH, Dugan CB, and Randles CI (1969) Microbial sulfate reduction and its potential utility as an acid mine water pollution abatement procedure. *Applied Microbiology* 17: 297–302.

Urrutia MM and Beveridge TJ (1994) Formation of fine-grained silicate minerals and metal precipitates by a bacterial surface (*Bacillus subtilis*) and the implications in the global cycling of silicon. *Chemical Geology* 116: 261–280.

van Gemerden H (1986) Production of elemental sulphur by green and purple sulphur bacteria. *Archives of Microbiology* 146: 52–56.

Wächtershäuser G (1988) Pyrite formation, the first energy source for life: A hypothesis. *Systematic Applied Microbiology* 10: 207–210.

Walsh MM and Lowe D (1999) Modes of accumulation of carbonaceous matter in the early Archaean: A petrographic and geochemical study of carbonaceous cherts from the Swaziland Supergroup. In: Lowe DR and Byerly GR (eds.) *Geologic Evolution of the Barberton Greenstone Belt, South Africa*, *Geological Society of America Special Paper 329,* pp. 115–132.

Wanger G, Onstott TC, and Southam G (2006) Structural and chemical characterization of a natural fracture surface from 2.8 kilometers below land surface: Biofilms in the deep subsurface. *Geomicrobiology Journal* 23: 443–452.

Westall F, Cavalazzi B, Lemelle L, et al. (2011) Implications of in situ calcification for photosynthesis in a ~3.3 Ga-old microbial biofilm from the Barberton greenstone belt, South Africa. *Earth and Planetary Science Letters* 310: 468–479.

Westall F, Boni L, and Guerzoni ME (1995) The experimental silicification of microbes. *Palaeontology* 38: 495–528.

Westall F, de Ronde CEJ, Southam G, et al. (2006a) Implications of a 3.472-3.333 Ga-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental conditions on the early Earth. *Philosophical Transactions of the Royal Society B* 361: 1857–1875.

Westall F, de Vries ST, Nijman W, et al. (2006b) The 3.466 Ga Kitty's Gap Chert, an Early Archaean microbial ecosystem. In: Reimold WU and Gibson R (eds.) *Processes on the Early Earth*, *Geological Society of America Special Paper 405,* pp. 105–131.

Westall F, Loizeau D, Foucher F, et al. (2013) Habitability on Mars from a microbial point of view. *Astrobiology* 13: 887–897.

Westall F and Southam G (2006) Early life on Earth. *American Geophysical Union, the Geophysical Monograph* 164: 283–304.

Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS* 87: 4576–4579.

Youssef MI (1965) Genesis of bedded phosphates. *Economic Geology* 60: 590–600.

Zahnle KJ, Freedman RS, and Catling DC (2011) Is there methane on Mars? *Icarus* 212: 493–503. http://dx.doi.org/10.1016/j.icarus.2010.11.027.

Zahnle K, Schaefer L, and Fegley B (2010) Earth's earliest atmospheres. *Cold Spring Harbor Perspectives in Biology* 10: a004895.

Zaug AJ and Cech TR (1986) The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* 231: 470–475.

10.15 Asteroids and Comets

YR Fernández, University of Central Florida, Orlando, FL, USA
J-Y Li, Planetary Science Institute, Tucson, AZ, USA
ES Howell, Arecibo Observatory/USRA, Arecibo, PR, USA
LM Woodney, California State University – San Bernardino, San Bernardino, CA, USA

© 2015 Elsevier B.V. All rights reserved.

10.15.1 Introduction 487
10.15.1.1 The Basics 487
10.15.1.2 Definitions 488
10.15.1.3 Dynamic Relationships 490
10.15.1.4 Relevance to Geology 490
10.15.1.5 Relevance to Meteorites 490
10.15.1.6 Relevance to Astronomy 492
10.15.2 Origins 492
10.15.2.1 Basics 492
10.15.2.2 Recent Models 493
10.15.2.3 Size Distributions 495
10.15.3 Surfaces 496
10.15.3.1 Background 496
10.15.3.2 Surface Morphology on Asteroids 496
10.15.3.2.1 Craters 496
10.15.3.2.2 Boulders 500
10.15.3.2.3 Lineaments 500
10.15.3.2.4 Smooth deposits 502
10.15.3.2.5 Regolith 502
10.15.3.3 Space Weathering 503
10.15.3.4 Surface Morphology on Cometary Nuclei 505
10.15.4 Structure, Interior, and Spin 508
10.15.4.1 Structure of Asteroids 508
10.15.4.2 Spin of Asteroids 509
10.15.4.3 Binaries 510
10.15.4.4 Cometary Properties 511
10.15.5 Composition 513
10.15.5.1 Basic Premises 513
10.15.5.2 Volatiles 514
10.15.5.3 Mineralogy 519
10.15.6 Big Questions and the Future 520
10.15.6.1 Where Is the Water? 520
10.15.6.2 Where Are the Organics? 520
10.15.6.3 What Was the Thermal and Compositional Gradient in the Protoplanetary Disk? How Much Mixing Was There? 521
10.15.6.4 What Did Small Bodies Do to Earth? 521
10.15.6.5 Forthcoming Breakthroughs 522
10.15.6.6 Longer-Term Missions 522
Acknowledgments 523
References 523

10.15.1 Introduction

10.15.1.1 The Basics

The importance of small bodies – asteroids, comets, and small satellites – to the overall picture of planetary formation and planetary evolution has been recognized for decades (see reviews by, e.g., Bottke et al., 2002b; Festou et al., 2004). Only since the 1990s, however, we learned enough about the small bodies and about solar systems in general to start to make detailed connections. Our advances have come mainly thanks to new opportunities in ground- and space-based telescopic observations of both our own solar system and other systems. Additionally, the multitude of spacecraft that have now visited a variety of small bodies have provided a wealth of detailed context. The 1990s, 2000s, and 2010s have brought incredible insight into the nature of the small bodies.

In this chapter, we will give a review of recent developments in our astronomical understanding of small bodies. To keep

the discussion manageable, note that we focus here mainly on comets and asteroids that are not in orbit around major planets; that is, we are excluding those small satellites that are likely captured comets and asteroids. Such satellites can inform and have informed our understanding of the general comet and asteroid population (see reviews by, e.g., Bell et al., 1993; Jewitt and Haghighipour, 2007; Turrini et al., 2009).

With this restriction, note that when we refer to 'small bodies' in this chapter, we will mean just comets and asteroids. In Section 10.15.1, we will discuss the historical and current definitions of small bodies and motivate their study by describing their scientific importance. In Section 10.15.2, we will describe the bodies' origins, their formation scenarios, their interaction with planets and with each other, and some of their dynamic processes. In Section 10.15.3, we discuss the nature of their surfaces, while in Section 10.15.4, we cover their interiors. Section 10.15.5 treats the composition of small bodies, and we finish in Section 10.15.6 with a summary of some of the big questions in and future directions of small-body research.

10.15.1.2 Definitions

Comets have been noted in recorded history for millennia, and asteroids have been known for over 200 years. For most of their observational history, the two groups of small bodies were thought to be clearly separate. More specifically, observational differences were assumed to translate directly to differences in the fundamental nature of the objects. Asteroids were point-like objects in the sky, mostly physically located between Mars and Jupiter in the asteroid belt (a.k.a. the main asteroid belt or just the main belt). Comets were extended sources – showing comae and tails that are generally much brighter than the solid-body source itself – and were either on parabolic or highly elliptical orbits that took them beyond the planetary region of the solar system or on short-period orbits that barely (if at all) went past Jupiter. Even the names given to the bodies revealed this dichotomy: asteroids were 'minor planets,' basically just smaller, fainter versions of the other planets that were seen to wander across the sky. Comets were entirely different beasts, historically not even recognized as astronomical phenomena (as opposed to atmospheric) until Tycho Brahe's parallax studies in the sixteenth century. Even the physical explanation for why comets looked extended – that is, that they had dust – was not established until the nineteenth century by, for example, Bessel (1836) and Bredichin (1877). The early twentieth century saw further characterization of the composition of the two groups, which further reinforced the apparent differences. Cometary spectra showed emission lines from gas species on top of a continuum spectrum, while those of asteroids sometimes showed broad absorption features in the continuum.

One of the major advances since the early 1990s is the realization that comets and asteroids are in fact not clearly separate classes of objects but instead part of a continuum of composition (e.g., Weissman et al., 1989, 2002). Comets are relatively icy and asteroids relatively rocky, but there are many objects that are now recognized to populate everywhere in between. For example, objects in the trans-Neptunian region – nominally called asteroids because of their lack of activity – are icy and would show cometary activity if they were only closer to the Sun (e.g., Barucci et al., 2002; Prialnik et al., 2008); objects that have orbits that are clearly cometary in nature are seen to be inactive, suggesting that their volatile component is deeply buried or now absent (e.g., Ishiguro et al., 2011); and sustained activity that has been seen to develop from objects in the asteroid belt between Mars and Jupiter appears to be driven by volatiles, indicating that ice must still exist somewhere on such objects that have been in the same thermal environment for billions of years (e.g., Hsieh et al., 2010).

Hence, while the terms 'asteroid' and 'comet' are still commonly used in the small-body community, there is broad recognition that these terms no longer necessarily act as shorthand descriptors for their true physical nature. In this chapter, we will continue to follow common usage when talking about specific objects that are referred to as comets or asteroids, but it is important to recognize that such terms can occasionally be misleading.

For this chapter, we will limit ourselves to a discussion of objects within a range of diameters. At the low end, we will not be talking about dust in and of itself but only dust as it relates to macroscopic comets and asteroids. In this sense, our lower limit to the diameter is about 0.1 m, although in practice, the lower limit is a few meters since almost no small bodies on heliocentric orbits have been observed that are smaller than that. At the upper end, we will limit ourselves to objects that have not pulled themselves into a roughly spherical or spheroidal shape as a result of self-gravity. This is consistent with the IAU definitions of planets and dwarf planets (IAU, 2006); we exclude these bodies from consideration here. We will, however, discuss (4) Vesta as it pertains to asteroid phenomena, though it arguably could be called a dwarf planet and has many terrestrial planet-like characteristics (e.g., Keil, 2002; Zuber et al., 2011).

Finally, a discussion of terminology is in order. Among comets, short- and long-period comets (SPCs and LPCs) are those that have orbital periods less than and longer than 200 years, respectively. SPCs include those of the 'Jupiter-family' (JFCs) and of the 'Halley-type' (HTCs), the former referring to comets whose orbits tend to be strongly influenced by Jupiter, have aphelia near Jupiter's orbit, are mainly in the ecliptic plane, and generally have orbital periods of about 6 years. HTCs are now thought to be dynamically evolved counterparts to the LPCs; they can have any inclination. This terminology is the traditional one; Levison (1996) discussed comet classification and his newer system is in use by some comet scientists. That convention divides the population into 'ecliptic comets' (ECs) and 'nearly isotropic comets' (NICs). ECs include not only the Jupiter-family but also a few other inner-solar system comets that have been decoupled from Jupiter. Furthermore, ECs can include the active 'centaurs,' which are objects orbiting among the outer planets beyond 5 AU and are in transition from the trans-Neptunian region to the inner solar system. NICs include the HTCs and LPCs and so dispense with the arbitrary 200-year boundary.

Among asteroids, most of the over 380,000 numbered asteroids known reside in the asteroid belt between Mars and Jupiter. Asteroids are assigned permanent numbers when their orbits are sufficiently well determined to be identified decades in the future without ambiguity. An additional 260,000 unnumbered asteroids are currently cataloged and may become numbered as new observations are made, and their orbits are better determined. Hundreds more asteroids are discovered in a typical month. Some objects are Mars crossers (MCs), coming closer to the Sun than Mars, and some objects

are Earth crossers, coming closer to the Sun than Earth. Some asteroids cross the orbits of Venus and Mercury as well. A near-Earth asteroid (NEA) is any object with a perihelion distance less than 1.3 AU, and the NEAs are broken down into 'Amors,' 'Apollos,' 'Atens,' and 'Atiras.' Amors are not Earth crossers, staying outside the Earth's orbit; Apollos and Atens do cross the Earth's orbit, with the distinction being whether the orbital semimajor axis is larger or small than 1 AU; Atiras also do not cross the Earth's orbit, staying entirely interior to it. While these definitions are still in use, it is important to note that many of these distinctions are no longer meaningful, as we now know that objects can easily move from one dynamic group to another due to both gravitational perturbations and non-gravitational forces.

A traditional dynamic distinction between comets and asteroids is the Tisserand parameter (T_J). In a solar system consisting of just the Sun, Jupiter on a circular orbit, and the small body (called the 'restricted three-body problem'), this parameter is nearly constant (i.e., an 'integral of the motion' in classical mechanics parlance). In the real solar system, it is still nearly constant and useful since in general asteroids have $T_J > 3$ and comets have $T_J < 3$. This comes from the fact that SPCs tend to be dynamically coupled to Jupiter; the Tisserand parameter can be related to the encounter velocity of a small body with Jupiter. However, there are many asteroids with $T_J < 3$ and several comets with $T_J > 3$, so the distinction is not exact.

Some of the classification distinctions can be seen in Figure 1, which shows a plot of the semimajor axis versus eccentricity of many of the asteroids in the inner solar system. In particular, the curves 'EC,' 'MC,' and 'JC' indicate the curves that mark which asteroids can cross the orbits of Earth, Mars, and Jupiter, respectively. We will return to this

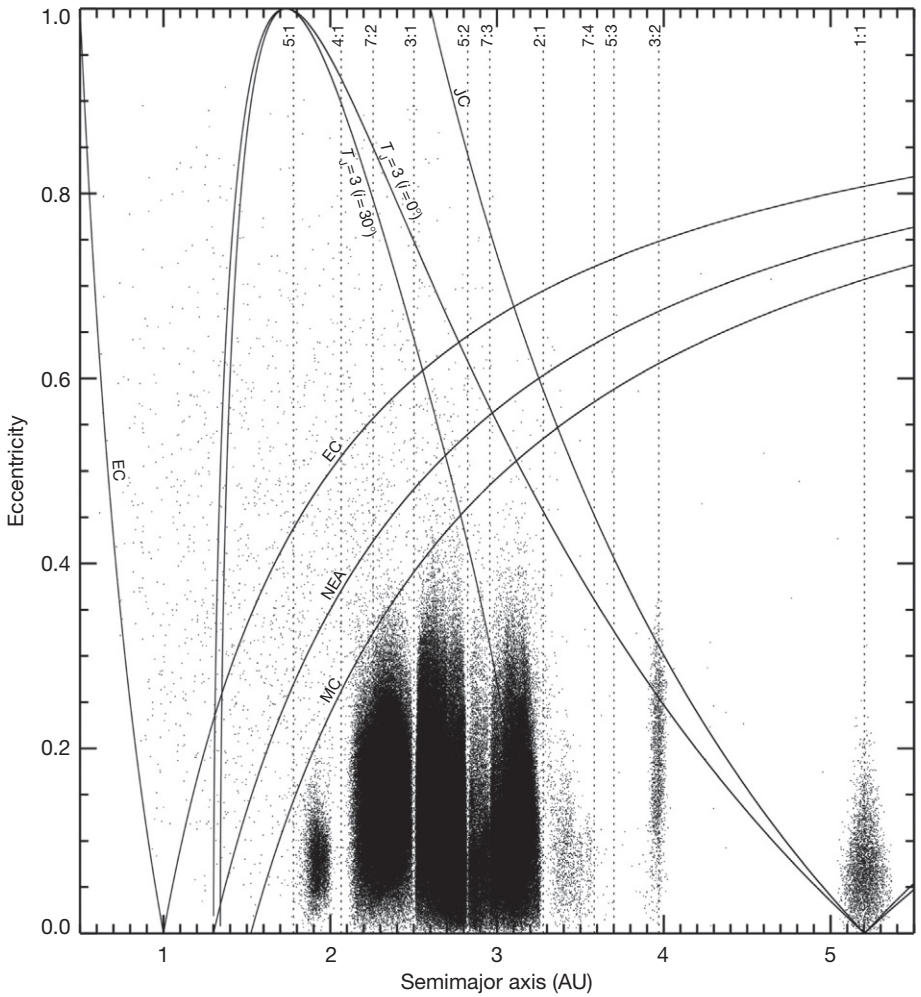


Figure 1 Orbital elements of the over 363,000 numbered asteroids (at time of writing), as represented in *a*–*e* space, where *a* is the semimajor axis in AU and *e* is the orbital eccentricity. Note that 1 AU is currently defined as 149,597,870.7 km exactly and is very nearly identical to the average Sun–Earth distance. There is obvious structure in the asteroid belt, and there are regions of orbital element space that are favored. 'NEA' indicates the locus where perihelion distance *q* is 1.3 AU, the defined boundary for near-Earth objects. 'EC,' 'MC,' and 'JC' indicate the Earth-, Mars-, and Jupiter-crossing boundaries, respectively. The two curves marked '$T_J = 3$' show where the Tisserand parameter is 3.0 for two possible values of the orbital inclination. Vertical dotted lines indicate mean-motion resonances (MMRs) with Jupiter; the resonance is listed at the top of each line. Note that Jupiter's semimajor axis is 5.20 AU; Mars's is 1.52 AU.

figure in later sections when we discuss various dynamic properties of the small bodies.

Note that in this chapter, we will give the complete designation when referring to an asteroid or comet for the first time, but references thereafter will use only the name (unless there is potential cause for confusion).

10.15.1.3 Dynamic Relationships

Some of the subpopulations of the solar system's small bodies are dynamically linked. In particular, the JFCs are linked to several other groups. The 'scattered disk' of the Kuiper Belt – containing objects beyond Neptune on high-eccentricity orbits – is thought to be the original source of these comets, and the objects progress in a dynamic cascade from beyond Neptune to the outer-planet region, essentially being handed off from the control of Neptune, to Uranus, to Saturn, and finally to Jupiter (Duncan et al., 2004). Jupiter-controlled comets have aphelia near 5 AU, and these are the JFCs we see today. As mentioned earlier, many members of the transitional population, the centaurs, have been discovered inhabiting the outer-planet region, and some of the centaurs show cometary activity. The timescale for an object to exist among the giant planets is only a few million years (Horner et al., 2004; Tiscareno and Malhotra, 2003).

JFCs share some of their orbital parameter space with NEAs. Current work suggests that up to a few percent of the NEA population could be JFCs that have simply lost their surface volatiles (e.g., Bottke et al., 2002a; DeMeo and Binzel, 2008; Fernández et al., 2005).

The connection between Jupiter's Trojan asteroids – objects in 1:1 resonance, but on average about 60° ahead or behind Jupiter in orbital longitude – and the JFCs is uncertain. It is possible that there is some leakage from the Trojan swarms to the comet population (Marzari and Scholl, 2002); however, no one has ever reported an observation of a Trojan having a comet-like coma or tail. Nor is it clear what would be an observational difference between a JFC that originated in the Trojan swarm and one that originated in the Kuiper Belt. One possible distinction is that Trojans might be expected to have a lower D/H ratio than objects that formed in the outer solar system, depending on how much compositional radial mixing has occurred; we discuss D/H ratios in more detail in **Section 10.15.5**.

The existence of a link between the asteroid belt and the NEAs is well established, although the details of which dynamic process moves which asteroids are still under debate (e.g., Bottke et al., 2002a).

10.15.1.4 Relevance to Geology

The most obvious geophysical link between Earth and the small bodies is cratering. Many craters are visible around Earth today, and the only reason we do not see more of them is because the Earth's surface is in constant change. Most – though not all! – asteroids have numerous craters, and indeed, most of the larger asteroid surfaces are saturated with craters. The study of the abundant craters on other solar system objects can bring insight into the impact process and how an impact's effects depend on the composition, density, and other bulk parameters of both the impactor and the target body.

Erosive processes that garden and turn over surfaces occur on most, if not all, small bodies. The main difference is that small bodies are airless bodies – none have an atmosphere in the same sense that Mars, Venus, and Earth (or even Mercury and the Moon for that matter) have atmospheres. The only thing that comes close is the gravitationally bound coma that sometimes appears around the largest cometary nuclei (e.g., Meech et al., 1997). But in general, no asteroid that we consider here has an atmosphere, and comets only have exospheres. The building up or tearing down of surface structures comes mainly from external effects (e.g., micrometeorite bombardment and impact cratering), though also perhaps from internal processes in the case of comets.

As will be discussed in later sections, the surfaces and interiors of small bodies can be quite different from those of Earth and the other terrestrial planets. Densities, porosities, compositions, collision histories, and surface processes can all be different from what we know on Earth. Yet all of these properties are crucial for understanding the geophysics of small bodies. Complicating our study of these objects is the fact that detailed geomorphology can only be obtained through the use of spacecraft or radar echoes, meaning that many comets and asteroids simply have not been studied in sufficient detail for there to be a true geophysical analysis. One important point to note is that spacecraft flybys of comets and asteroids have made it abundantly clear that we have not yet seen the full diversity of small bodies close-up.

10.15.1.5 Relevance to Meteorites

We know through laboratory or in situ sample analysis that only a small fraction of meteorites originate from the Moon and Mars; the vast majority are from asteroids. The meteorites thus give us important insights into the composition of the solar nebula from which the planets formed. A comprehensive review of meteorite studies is beyond the scope of this chapter and indeed is the subject of entire volumes (e.g., Lauretta and McSween, 2006). We rely heavily on the studies of meteorites to give us an understanding of the composition of asteroids while at the same time recognizing that the sampling of the asteroid belt may be extremely nonuniform. Attempts to link telescope spectra of asteroids with laboratory measurements of meteorites have shown that there is not a simple one-to-one correspondence and many asteroids do not seem to be represented in the meteorites we have. On the other hand, some links are considered very secure, such as the spectra of howardites, diogenites, and eucrites matching the spectrum of asteroid Vesta; these meteorites are considered to be samples of this asteroid (Binzel and Xu, 1993; Drake, 2001). The identifiable spectrum of Vesta has let us identify collisional fragments of previous impacts in the asteroid belt and among the NEAs (Binzel et al., 2004). However, other connections of specific asteroids to specific meteorites are not so clear.

Several asteroid taxonomies have been suggested to better understand the relationships between different asteroid spectral types, meteorites, and compositions. Tholen and Barucci (1989) gave a thorough review of the most commonly used taxonomic system based on photometry in the visible range

(0.35–1.0 μm). Bus and Binzel (2002) used higher-resolution spectra over a slightly narrower spectral range (0.45–0.95 μm) to extend the Tholen taxonomy to include more specific spectral features. DeMeo et al. (2009) extended the spectral range from visible to near-infrared (0.4–2.5 μm), and this Bus–DeMeo taxonomy is now the most widely used. Three groups of spectral types, S-, C-, and X-complex, encompass the majority of the asteroids. Each complex includes several related taxonomic classes designated by one or two letters; the specific taxonomic class that an asteroid is assigned depends on the details of the spectral shape. **Figure 2** summarizes the spectra of all the taxonomic classes in the Bus–DeMeo system; it shows both the average spectrum of each taxonomic type and some actual asteroid spectra. This figure shows the diversity of asteroid reflectances in visible and near-infrared wavelengths among all the taxonomic types. The S-complex asteroids are primarily pyroxene and olivine, and in some cases metallic iron. Ordinary chondrites (OCs) and other stony but largely undifferentiated meteorites are thought to be samples of S-complex asteroids. The C-complex asteroids are mostly low-albedo, flat featureless spectra except for a broad absorption band at 0.7 μm indicative of oxidized iron (Vilas and Gaffey, 1989) defining the Ch class. Various carbonaceous chondrite meteorites are thought to be samples of these types of asteroids, although the spectral matches are not perfect. The X-complex asteroids have been linked to many different meteorite groups such as aubrites, enstatite chondrites, and achondrites. Links to iron–nickel meteorites are problematic; originally, these were thought to come from the so-called (in the Tholen system) M-type asteroids, but now, it is unclear if any of the large asteroids can be primarily iron–nickel (see, e.g., Shepard et al., 2010, submitted for publication, and references therein for detailed discussion). A few additional taxonomic classes such as A, V, and R are unusual or match specific asteroids. For example, the V types are most likely related to Vesta, as discussed in the succeeding text. Finally, the D-type asteroids have featureless and very red sloped spectra and match the outer belt asteroids and Jupiter Trojans but do not match any meteorites very well, except perhaps Tagish Lake (Hiroi et al., 2001). Interestingly, both of Mars's moons are D-type objects; this is unusual since D types are not so abundant in the inner part of the solar system. For convenience in this chapter, and unless otherwise noted, we will refer to broad spectral groups of asteroids using the Bus–DeMeo taxonomy.

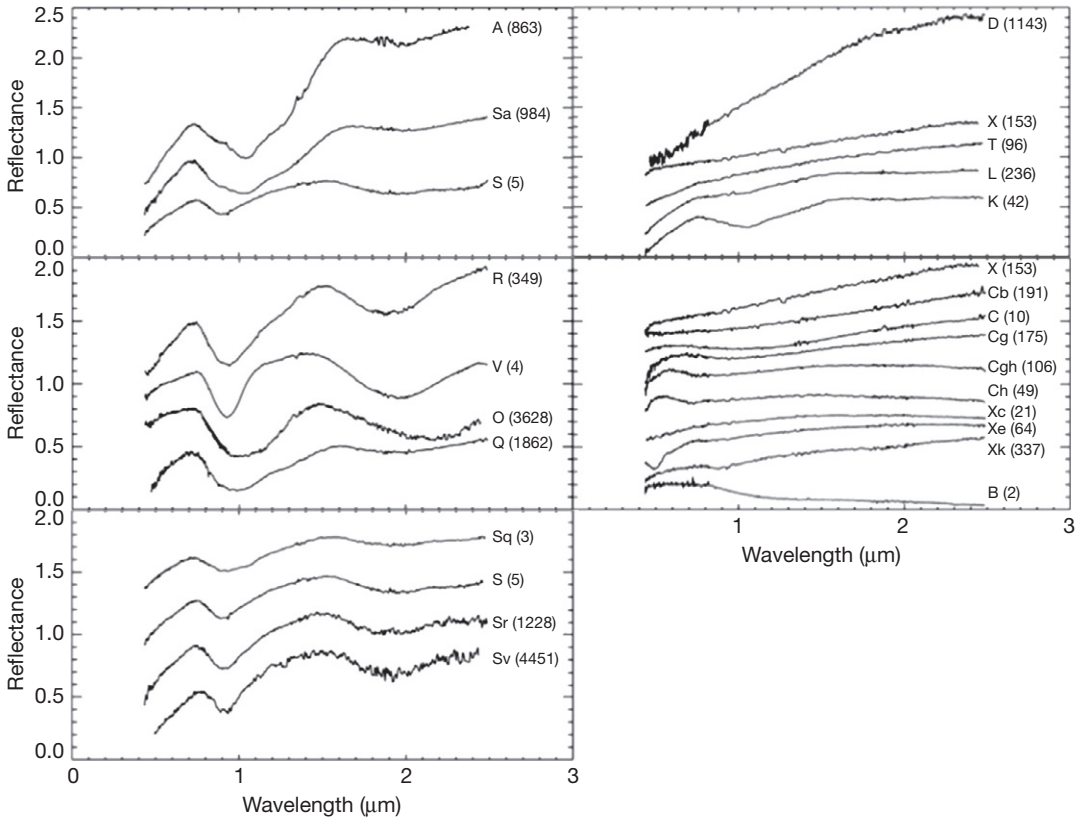


Figure 2 The range of asteroid relative reflectances as seen in the red and near-infrared. There are 26 spectra here, representing the 24 taxonomic classes for the near-infrared devised by DeMeo et al. ('S' and 'X' are shown twice); the taxonomic type is the letter at the right edge of each spectrum. The number after the type is the numbered asteroid that represents the prototype. Note that for clarity, spectra within a panel have been shifted vertically; reflectance is set at unity for a wavelength of 0.55 μm. Asteroids on the left all show the classic two-band ('Band I' and 'Band II') absorption features near 1 and 2 μm due to pyroxene and olivine. These types include the objects in the S-complex (S, Sa, Sq, Sr, and Sv). Asteroids on the right lack these absorptions, including objects in the C-complex (C, Cb, Cg, Cgh, Ch, and B) and X-complex (X, Xc, Xe, and Xk); indeed, many of the asteroids on the right lack distinguishing diagnostic absorption features, making determining their composition problematic. Reprinted from DeMeo FE, Binzel RP, Slivan SM, and Bus SJ (2009) An extension of the Bus asteroid taxonomy into the near-infrared. *Icarus* 202: 160–180, Figures 5–7, 11, and 14, with permission from Elsevier.

The connection between meteorites and comets is very sparse; no meteorite currently in hand is considered to have definitely come from a comet. Cometary meteorites are expected to be rare simply due to high entry speed (Campins and Swindle, 1998), although it is possible that some relatively slow meteor showers may provide meteorites that survive (Brown et al., 2013).

10.15.1.6 Relevance to Astronomy

The asteroids and comets we see today are the survivors of the planetary formation process that proceeded during the start of our solar system. The current dominant hypothesis is that the nebular cloud that formed the Sun also had a protoplanetary disk, from which the first planetesimals, and then planets, formed. The details of what the protoplanetary disk was like and the actual mechanical process that formed the planets are both major questions in astronomy that are the subject of significant research. Thus, the study of comets and asteroids is tied to some of the 'holy grails' of astronomy. The role of comets and asteroids in the history of our solar system's planets has become even more important since the late 1990s and 2000s with the discovery of many other planetary systems in our galaxy.

One of the interesting results from the several hundred exoplanets that have been discovered is the fact that the architecture of our own solar system seems to be relatively unusual (e.g., Borucki et al., 2011; Udry and Santos, 2007). Our planetary system's layout – with small terrestrial planets close to the home star and large gas- and ice-giant planets farther out – has not yet been seen in very many other systems. While observational biases and limitations have made it difficult to ascertain just how many terrestrial planets there are elsewhere, it is clear that there are a great many ways for giant planets to end up being distributed within a planetary system. This reality has motivated dynamicists to investigate possible modes of planetary migration. Some of this migration is influenced by the population of small bodies; essentially, the act of perturbing small bodies may transfer sufficient angular momentum (by the law of conservation) to change a planet's orbit. Thus, the population of small bodies in the protoplanetary disk is a fundamental part of understanding the development of the planets' final layout in the solar system. Furthermore, we are discovering other systems around older stars where clearly there are asteroid or comet belts with masses much larger than those we currently have in our solar system (e.g., Lawler et al., 2009; Lisse et al., 2007). These disks of material are not protoplanetary, since the parent stars are usually hundreds of millions or billions of years old. Two questions stemming from these observations are why some systems have such massive disks and whether our own solar system went through a similar stage in the past.

Another area of relevance that pertains to the solar system itself has to do with water and organic molecules. The existence and distribution of these substances are, of course, crucial things to understand since life on Earth (and so far, life as we know it) requires such compounds to exist. Both comets and asteroids contain water and organics in various forms and concentrations. Current thinking is that it is unlikely that Earth formed with its current water content, since it was likely too hot; most of the Earth's water presumably was brought to it via small-body impacts (e.g., Drake and Campins, 2006; Mottl et al., 2007). The original organic component of Earth was likewise probably miniscule, with small bodies perhaps providing the bulk of the organics (Anders, 1989).

Lastly, small bodies have relevance today since they represent an impact hazard to Earth. The extreme case is the K–Pg mass extinction 65 million years ago, after which a large fraction of the Earth's species became extinct in a geologically short time. Presumably, an important factor in instigating this extinction was the impact of an approximately 10 km wide object striking what is now the Yucatán Peninsula (Alvarez et al., 1980). Many other impacts have happened since that event, causing varying degrees of damage to local and regional ecosystems and in some cases whole biomes. A very recent and well-observed event was the arrival of the ~15 m wide Chelyabinsk bolide in February 2013. That asteroid, which was not known before the encounter, was possibly the largest object to strike the Earth's atmosphere since the 1908 Tunguska event (Brown, 2013). Somewhere out there is the next devastating small body that will eventually collide with us; the questions are how big is it, and is it coming in a few years, centuries, millennia, or longer? Unlike the dinosaurs, we can actually attempt to answer this question and try to guard against catastrophe.

10.15.2 Origins

10.15.2.1 Basics

The traditional, fundamental compositional distinction between comets and asteroids is the ice content. The zeroth-order explanation for that distinction involves the existence of a 'snow line,' a distance from the forming Sun within the protoplanetary disk beyond which water could condense. While the term 'snow line' conjures up the idea of a sharp boundary in the protoplanetary disk, it appears (as mentioned in **Section 10.15.1**) that the reality is more complicated, with the disk having a trend of ice abundance that varied with time rather than a sharp falloff (see, e.g., the review by Encrenaz, 2008). The most widely accepted idea is that the comets and asteroids we see today are the remnants of the planetesimals that formed by accretion within the protoplanetary disk.

The specifics of the accretion process still provoke a major ongoing research effort, most strikingly epitomized by the so-called meter barrier. While numerical modeling has been able to reproduce the accretion of solid-phase rock, metal, and ice grains into conglomerates up to submeter scales with the correct timescales, simulating the creation of larger bodies has been difficult. The main problem is that, in many simulations, a two-body collision between meter-size objects seems to result in a destructive collision rather than an accretion. One possibility is that the interaction between the meter-size boulders and turbulent gas in the disk can sufficiently affect the nature of the collisions so that kilometer-scale bodies can actually be created (e.g., Chiang and Youdin, 2010; Johansen et al., 2007).

There is evidence for a protoplanetary disk within which species aside from ice exhibited trends with distance from the Sun. The distribution of spectroscopic asteroid types is not randomly mixed, but instead follows a trend with heliocentric distance, which has been interpreted as a remnant of the solar nebula. This phenomenon has been known for decades (e.g.,

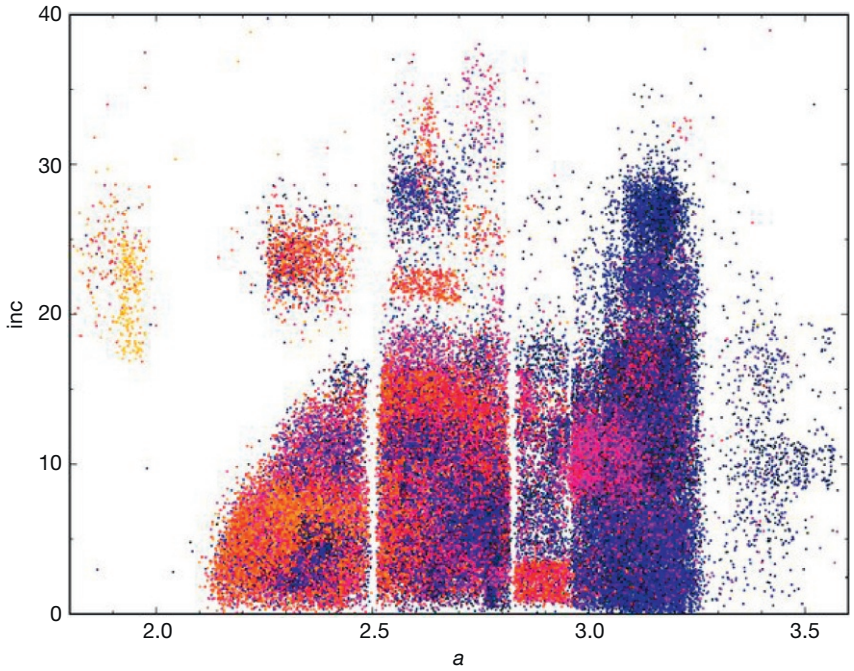


Figure 3 Scatter plot of main-belt asteroids in *a*–*i* space (where *i* is the orbital inclination) showing the albedos of the asteroids as derived by observations with the WISE spacecraft. The color coding of the points gives the albedo, with black and gray as the lowest albedos (0.05 and below), increasing to blue, purple, pink, red, and orange, with yellow showing the highest albedos (above 0.35). There is a clear trend of albedo with distance from the Sun; the trend exists for composition as well. Reproduced from Masiero JR, Mainzer AK, Grav T, et al. (2011) Main belt asteroids with WISE/NEOWISE. I. Preliminary albedos and diameters. *Astrophysical Journal* 741: 68, Figure 12, ©AAS, with permission.

Gradie et al., 1989), but a recent demonstration of the situation is shown in **Figure 3**. Thousands of asteroids observed by the WISE spacecraft (Masiero et al., 2011) show that asteroid albedos (and similarly, taxonomic types and compositions) are not uniformly distributed in the asteroid belt; more primitive (lower albedo) objects dominate farther out, while more processed (higher albedo) objects dominate closer in. The correlation of spectral types with distance motivates the idea that the asteroid belt can be used to better understand the conditions in the solar nebula and the material that formed the terrestrial planets. However, discoveries of extrasolar planets and other planetary systems have greatly broadened our notions of what is 'normal' and 'typical.' Also, increasing computation capabilities have allowed ever more realistic and detailed models of accretion and are now challenging long-held ideas of what we thought we knew about solar system formation. The asteroid belt and Kuiper Belt are important keys to constraining and testing these models.

The mass of the asteroids in the asteroid belt between Mars and Jupiter is about 6×10^{-4} times the mass of Earth, extremely depleted compared to a protoplanetary disk extending smoothly from 1 to 5 AU and likely originally containing many Earth masses of material (e.g., Morbidelli et al., 2012). The formation of Jupiter was long thought to be responsible for disrupting the formation of a planet in this region and for scattering material around the solar system. The asteroids remaining in the asteroid belt have dynamic signatures from Jupiter's gravitational effect, such as the Kirkwood gaps – evident as the gaps in semimajor axis as seen in **Figure 1**. These gaps are created by mean-motion resonances (MMRs) with Jupiter, where the orbital period of the asteroid is a small-integer ratio of Jupiter's; they are denoted at the top of **Figure 1**. An MMR tends to enhance perturbations from Jupiter and so helps kick the asteroid out of that region of orbital element space. The belt's inner edge is controlled partially by Mars. This is also seen in **Figure 1**; most of the asteroids now in the asteroid belt do not cross the MC line. The belt is also sculpted by the ν_6 resonance – a secular resonance arising from when the asteroid's longitude of perihelion has a precession frequency matching that of Saturn (see, e.g., Froeschlé and Morbidelli, 1994). This resonance is most noticeable in a plot of inclination versus semimajor axis, as in **Figure 3**; the reason the asteroids look confined to low inclinations is because the ν_6 resonance tends to kick out any asteroid with an inclination that is too high. The equivalent resonance with Jupiter, ν_5, operates at higher inclinations, so in this case, Saturn's influence is stronger than Jupiter's.

10.15.2.2 Recent Models

Recent modeling efforts to reconcile formation timescales of the outer planets have led to consideration of large-scale migration of the giant planets. The so-called Nice model (Gomes et al., 2005; Morbidelli, 2010; Morbidelli et al., 2005; Tsiganis et al., 2005) – named after the city in France where many of the model's originators work – has come to be broadly accepted, although the details are still debated. Indeed, the model, while explaining many observed orbital properties of small bodies in our solar system, has only made limited observational

predictions and could not account for the small mass of Mars (compared with Earth and Venus).

The basic premise of the model is that the outer planets (Jupiter, Saturn, Uranus, and Neptune) started out in orbits much closer to the Sun than they are now. There was also a disk of material – forerunners to today's comets – that was the remnant of the protoplanetary disk from which the four planets formed. Scattering of these small bodies (as mentioned in Section 10.15.1) caused the planets to migrate in semimajor axis, eccentricity, and inclination. Approximately 800–900 million years after the planets formed, Saturn and Jupiter happened to fall into a 2:1 MMR. This had a profound effect on the layout of the solar system, strongly perturbing the remaining small bodies in the outer solar system, as well as Uranus and Neptune themselves. There is even a possibility that Uranus and Neptune switched order; this falls out of some simulations and would nicely explain the long-standing mystery of why Neptune is more massive than the next planet closer in to the Sun. An example of the effects of the model is shown in Figure 4; small bodies in the vicinity are very rapidly scattered. The model can explain such other observables as the inclination distribution of the Jovian Trojan asteroids and the basic structure of the Kuiper Belt and would be the cause of the Late Heavy Bombardment seen in the lunar cratering record, about 3.9 billion years ago.

More recently, Walsh et al. (2011) had added an early, inward migration of Jupiter to 1.0 AU, which finally allows Mars to form with its small mass. However, this requires a complete reworking of our interpretations of the asteroid belt. The 'Grand Tack,' as this model has been called, is still being refined, but like the Nice model, it explains some long-standing mysteries in an appealing way. It too may be overturned, but at present, it provides a useful framework for understanding the asteroids that is substantially different from before. Morbidelli et al. (2012) gave a comprehensive review of the formation of the terrestrial planets, from nebular dust to present-day bodies. The asteroids are an important constraint, as the asteroid belt as seen today is not consistent with early Jupiter migration in the gas disk (Walsh et al., 2012) or the later outer-planet migration (jumping Jupiter) scenarios.

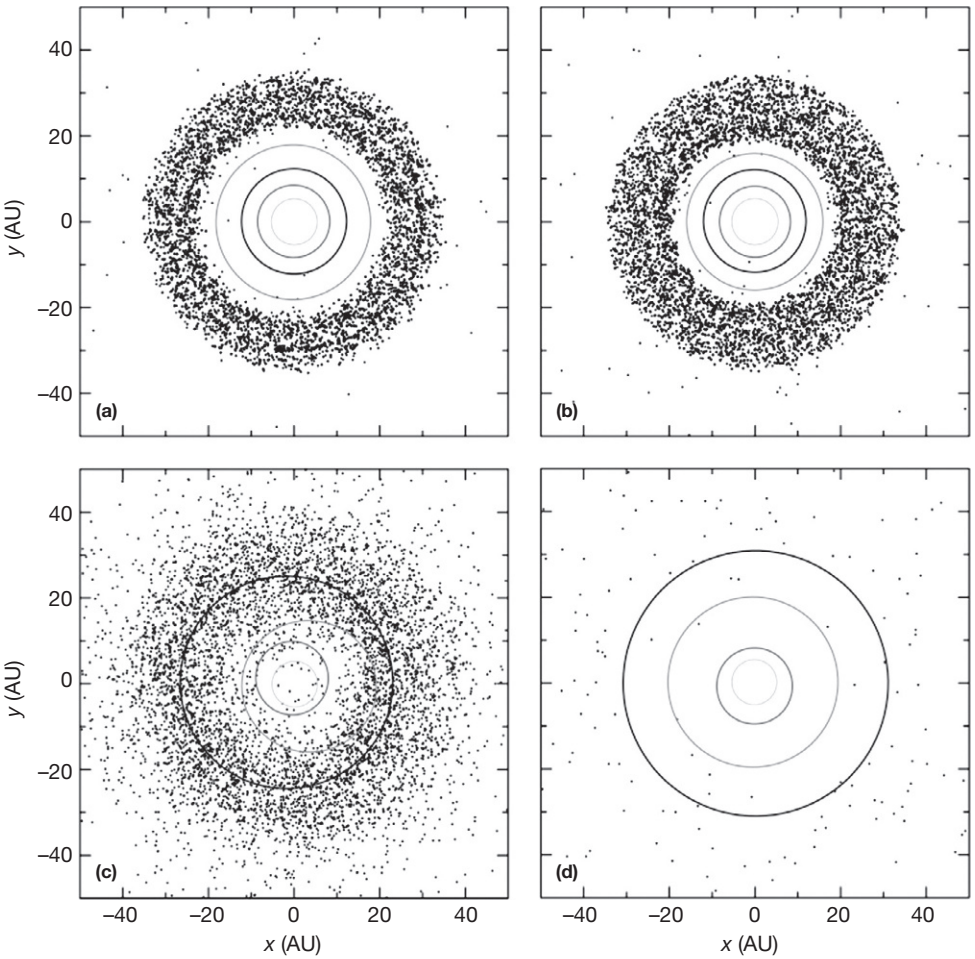


Figure 4 An example simulation of the Nice model. The four panels show the development of the four outer planets plus thousands of massless test particles through the use of a numerical simulation. Panel (a) represents the solar system 200 million years after the formation of the planets; panel (b), 879 million years; panel (c), 880 million years; and panel (d), 1080 million years. Jupiter and Saturn enter the 2:1 MMR at 880 million years, and the disk of test particles is quickly disrupted. Reprinted from Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469, Figure 2, with permission from Macmillan Publishers Ltd.

Finding dynamic models that can match the current inclination and eccentricity distribution of the largest asteroids is surprisingly difficult and serves as a sensitive measure of the dynamic evolution of the solar system.

Morbidelli et al. (2012) and references therein described how the original material in the asteroid belt would have been disrupted by inward migration of Jupiter, leaving only the inner asteroid belt intact and scattering some of these objects outward behind it as it moves. After Saturn moved into resonance with Jupiter, both began to migrate outward, and some of the scattered objects were scattered back into the asteroid belt region, along with many additional volatile-rich objects (C-type asteroids) that may have formed closer to 5 AU. This scenario solves the problem of having both high- and low-temperature materials side by side in the asteroid belt. However, it means that the compositional gradient with heliocentric distance is not primordial and allows a larger amount of radial mixing than is usually considered probable. The problem of linking meteorites to their asteroid parent bodies also becomes nearly impossible, since the current position of the asteroids may not be at all related to where those bodies formed. However, the observations that asteroid families, collisional fragments of once intact parent bodies, are nearly all spectrally uniform, while many meteorites clearly show that differentiation has occurred, is no longer an impossible situation. Differentiated objects can form closer to the Sun and be emplaced in the asteroid belt by Jupiter during migration. Undifferentiated objects form farther out and are scattered inward by Jupiter. Both can now exist side by side, but neither is representative of material that condensed at 2–4 AU. We may have little hope to identify primordial material, if this accretion model is correct.

The topic of the origin of the Kuiper Belt, which is likely the original source of the JFCs we see today, is discussed in detail in **Chapter 10.19**. Briefly, the orbital distribution of today's belt can be a useful constraint on the historical dynamic effects that sculpted it. For example, the slow outward migration of the outermost planet may have swept up some objects into MMRs; Pluto, residing in Neptune's 3:2 resonance, is the most famous example. But the belt also holds a 'scattered disk' of objects that were presumably scattered by Neptune into highly eccentric orbits. There is also a 'hot' classical disk that maintains relatively low eccentricity but is dynamically 'hot' in the sense of having higher inclination. In contrast, the 'cold' classical disk has lower-inclination objects. There seems to be some difference in origin of these various populations since there is some difference in their surface properties – namely, color (e.g., Romanishin et al., 2010; Sheppard, 2012), which is for the most part the only measurable property available for the vast majority of known Kuiper Belt objects. As mentioned earlier, the Nice model would suggest that Saturn's 2:1 MMR crossing disrupted the primordial Kuiper Belt, and it would imply that we may have a difficult time disentangling primordial and evolutionary effects.

10.15.2.3 Size Distributions

The size distribution of the asteroids has been measured to increasingly smaller sizes, but is still not linked to the dust distribution with any confidence. Models and observations have finally come together and agree that below about 10 km in diameter, all main-belt asteroids are collisional fragments (e.g., Bottke et al., 2005). Asteroids bigger than 30–50 km are likely to be fractured from multiple collisions but have lifetimes of 4.5 Gy or more and could be survivors of the formation of the planets. However, if the source regions for the present-day asteroid belt are as broad as 1—6 AU or more, these limits may change. There is some observational evidence for waviness in the cumulative size distribution – that is, for a distribution that follows a power law but with a bump at certain sizes (e.g., O'Brien and Greenberg, 2005). The cumulative size distribution as shown in the review by Bottke et al. (2005) is shown in **Figure 5**. This waviness could be interpreted as being controlled by the strength–size relationship of the asteroids. However, the existence of the waviness, and its location in the size distribution, is still under debate (e.g., Gladman et al., 2009).

The current size distribution of comets is, compared with the asteroids, poorly known. Work on the JFCs has been done by several teams (e.g., Fernández et al., 2013; Meech et al., 2004) and suggests that there is currently a lack of sub-kilometer-size comets. Some recent estimates of the distribution are shown in **Figure 6**. Note that there is no expectation that the current observed distribution of comet sizes reveals the primordial distribution; the collisional environment in the Kuiper Belt is likely too violent (Farinella and Davis, 1996). The observational situation is even worse for LPCs. Using the current size distribution of comets to work backward and tease out the original size distribution is a very long-term project that will require having a much better understanding of collisions, of cometary mass loss, and of fragmentation.

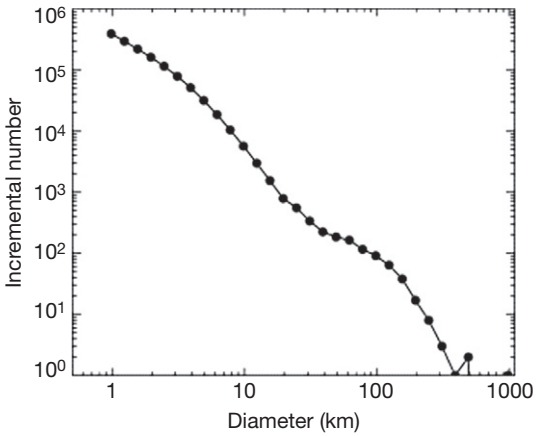


Figure 5 Main-belt asteroid size-frequency distribution, as shown in Figure 1 from the work of Bottke et al., 2005. This distribution is computed from the absolute magnitude H distribution, transformed into diameter bins assuming an average geometric albedo p_v of 0.092. For the fainter objects, those with $H > 12$ (i.e., a diameter less than 18 km for the average albedo), the data are from the Sloan Digital Sky Survey. For the brighter objects, with $H < 12$, the observational census is virtually complete and so the data are from the list of known asteroids. Reprinted from Bottke WF, Jr., Durda DD, Nesvorný D, et al. (2005) The fossilized size distribution of the main asteroid belt. *Icarus* 175: 111–140, with permission from Elsevier.

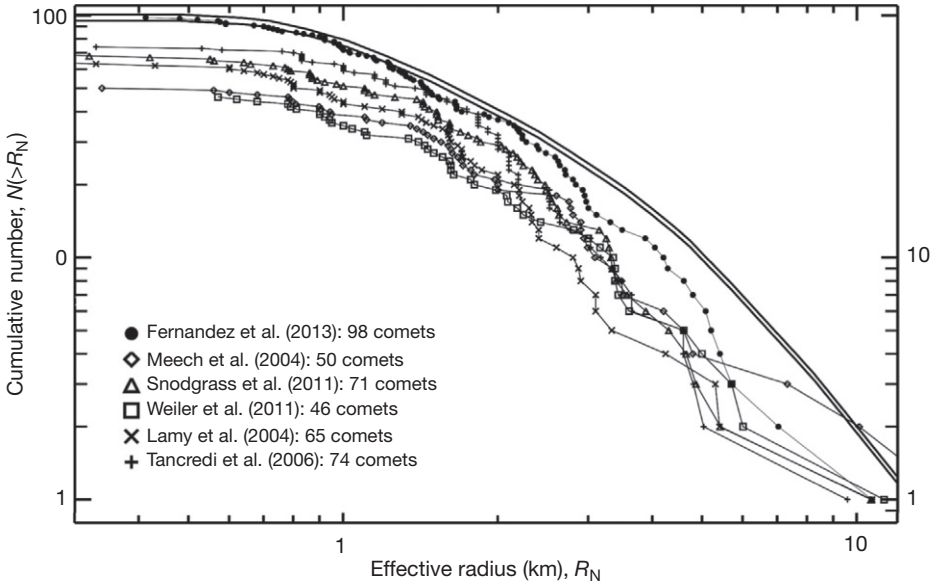


Figure 6 Cumulative size distributions of the Jupiter-family comets as reported by various workers. The legend indicates how many JFCs were used in each result. All results used primarily visible-wavelength measurements of nuclei, except for Fernández et al., who used exclusively infrared measurements. There is broad agreement in the distributions. The double-solid lines indicate the intrinsic distribution model by Meech et al. (2004) that best reproduces their observed distribution; this model requires that there be a lack of sub-kilometer JFCs to account for the small number of such comets that have been discovered. Adapted from Fernández YR, Kelley MS, Lamy PL, et al. (2013) Thermal properties, sizes, and size distribution of Jupiter-family cometary nuclei. *Icarus* 226: 1138–1170, with permission from Elsevier.

10.15.3 Surfaces

10.15.3.1 Background

Twelve asteroids and five comets have been imaged at close range by spacecraft through flybys and/or rendezvous, returning a huge number of disk-resolved images of these small bodies. Table 1 lists all small-body encounters within 10^6 km. Some views of these flyby targets are shown in Figures 7–10.

These planetary exploration missions have fundamentally changed solar system small bodies from astronomical objects to geologic objects. Our knowledge about the surface morphology and cratering characteristics has been greatly expanded as a result of those planetary missions. We will discuss the surface morphology of comets and asteroids separately; both kinds of objects are affected by cratering, but cometary surfaces are also changed on much shorter timescales by outgassing and erosion due to sublimation.

10.15.3.2 Surface Morphology on Asteroids

The first ever spacecraft flyby of an asteroid was in 1991, when the Galileo spacecraft visited (951) Gaspra while en route to Jupiter (Belton et al., 1992; Veverka et al., 1994). This flyby marks the start of observational studies of asteroid geomorphology. Since then, 11 more asteroids have been visited through either flybys or rendezvous missions (Table 1). Among those asteroids, (9969) Braille, (5535) Annefrank, and (132524) APL were observed at relatively low resolutions that do not allow for detailed studies of their surface morphology. The remaining asteroids represent a considerable sample of the general morphological properties of asteroids, from a small size of about half a kilometer in the maximum dimension for (25143) Itokawa (Demura et al., 2006) to the second largest of all main-belt asteroids, Vesta (Jaumann et al., 2012), and from near-Earth population – (433) Eros and Itokawa – to main-belt asteroids.

Despite their small sizes and therefore much weaker surface gravities compared with planets, asteroids display a large variety of surface features, including craters, boulders, lineaments (grooves, troughs, and ridges), rough and smooth terrains, regolith, and landslides. Being geologically inactive, the surfaces of asteroids are usually dominated by impact craters. The surface features reflect the mechanical responses of the target asteroids to various impact energies under various geometric settings and therefore reveal their physical properties. We summarize the major geologic features found on asteroids in this section, compare different asteroids, and discuss the implications to asteroids' evolutionary history and properties.

10.15.3.2.1 Craters

The dominant geomorphological features on the surfaces of almost all larger asteroids are craters, most of which are of impact origin, although this is not entirely certain for the small asteroid Itokawa due to the different morphology of the 'circular depressions' (Hirata et al., 2009). After formation, craters on asteroids are subject to size-dependent degradation and removal processes, such as regolith movement, slope failure, and micrometeorite bombardment. Three collective properties of crater populations on asteroids are the most important and widely studied. First, the spatial density is a critical indicator of surface age since the last resurfacing (e.g., Chapman, 2002; Ivanov et al., 2002). Second, the size-frequency distribution (SFD) is related to the SFD of impactors via cratering scaling laws (e.g., Holsapple et al., 2002, for a review; Holsapple and Housen, 2007; Melosh, 1989). The gravity and strength of asteroid material affect the scaling law, along

Table 1 List of small bodies visited in close proximity ($<10^6$ km) by spacecraft, listed in chronological order

Object	*Type*	*Year*	*Spacecraft*	*Closest pass (km)*	*Best res. (m/pixel)*	*Dimension of object (km)*
21P/Giacobini–Zinner	JFC	1985	ICE	7800	N/A	?
1P/Halley	HTC	1986	Vega 1	8889	~300	15×7
1P/Halley	HTC	1986	Vega 2	8030	~300	15×7
1P/Halley	HTC	1986	Giotto	596	130	15×7
1P/Halley	HTC	1986	Suisei	151,000	37,000	15×7
(951) Gaspra	S MBA	1991	Galileo	1600	54	$18 \times 11 \times 9$
26P/Grigg–Skjellerup	JFC	1992	Giotto	200	N/A	?
(243) Ida	S MBA	1993	Galileo	2400	25	$60 \times 25 \times 19$
(253) Mathilde	C MBA	1997	NEAR	1200	200	$66 \times 48 \times 46$
(433) Eros	S NEA	1998	NEAR	3827	363	$34 \times 11 \times 11$
(9969) Braille	Q NEA	1999	Deep Space 1	26	182	$2 \times 1 \times 1$
(433) Eros	S NEA	2000–2001	NEAR	0	cm to m	$34 \times 11 \times 11$
(5535) Annefrank	S NEA	2002	Stardust	3079	184	$7 \times 5 \times 3$
19P/Borrelly	JFC	2002	Deep Space 1	2171	47	8×2
81P/Wild 2	JFC	2004	Stardust	237	14	$6 \times 4 \times 3$
9P/Tempel 1	JFC	2005	Deep Impact flyby	700	10	8×5
9P/Tempel 1	JFC	2005	Deep Impact impactor	0	3	8×5
(25143) Itokawa	S NEA	2005	Hayabusa	0	cm to m	$0.5 \times 0.3 \times 0.2$
(132524) APL	S MBA	2006	New Horizons	101,867	~2000	2.3
(2867) Šteins	E MBA	2008	Rosetta	800	80	$7 \times 6 \times 5$
(21) Lutetia	C or M MBA	2010	Rosetta	3162	60	$121 \times 101 \times 75$
103P/Hartley 2	JFC	2010	Deep Impact	700	~7	0.7×2.3
9P/Tempel 1	JFC	2011	Stardust	181	12	8×5
(4) Vesta	V MBA	2011–2012	Dawn	210	20	$573 \times 557 \times 446$
(4179) Toutatis	S NEA	2012	Chang'e 2	3.2	10	$4.6 \times 2.3 \times 1.9$

Under 'Type,' 'HTC' refers to Halley-type comet; 'JFC,' Jupiter-family comet; 'MBA,' main-belt asteroid; 'NEA,' near-Earth asteroid. The letter before 'MBA' and 'NEA' refers to the taxonomic type; the types are discussed in **Sections 10.15.1.5** and **10.15.3.3**. Note that flybys of comets 21P/Giacobini–Zinner and 26P/Grigg–Skjellerup returned no imaging, hence we have put a '?' in the dimension column.

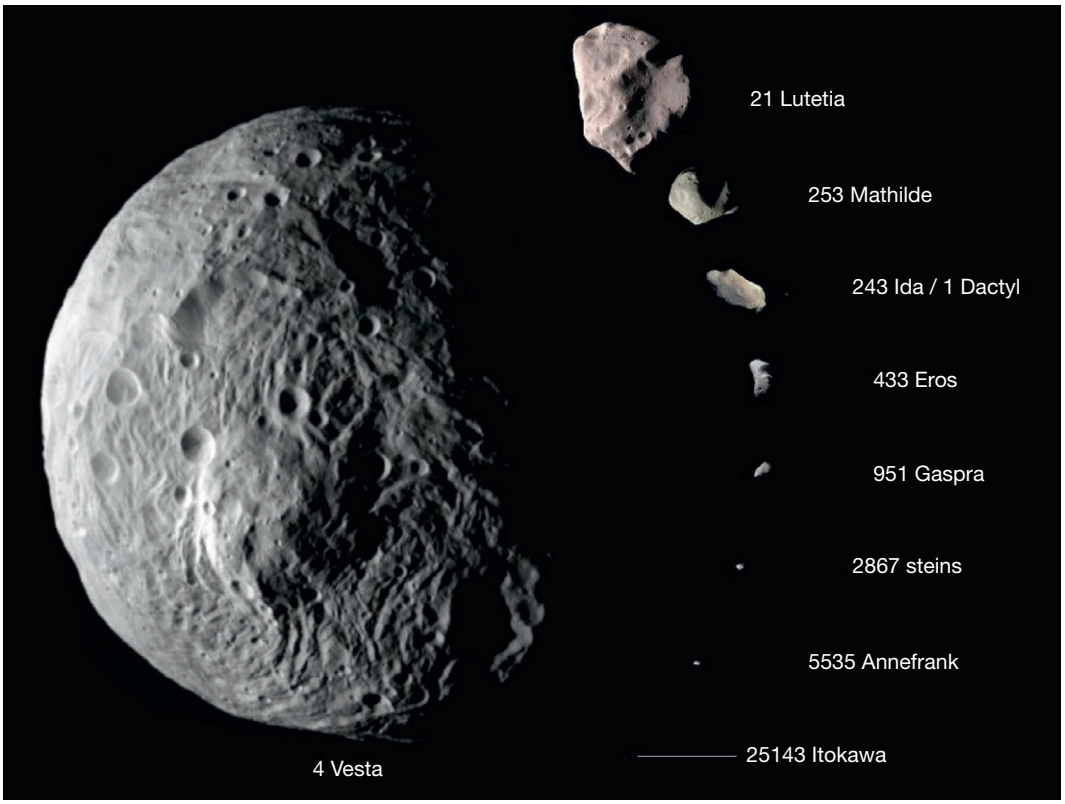


Figure 7 Montage of many of the asteroids that have had close spacecraft flybys, shown to scale. Note that the diameter of the largest object, Vesta, in this view is about 560 km, while that of the smallest object, Itokawa, is under 1 km. Courtesy NASA/JPL-Caltech/JAXA/ESA (Photojournal image PIA14316).

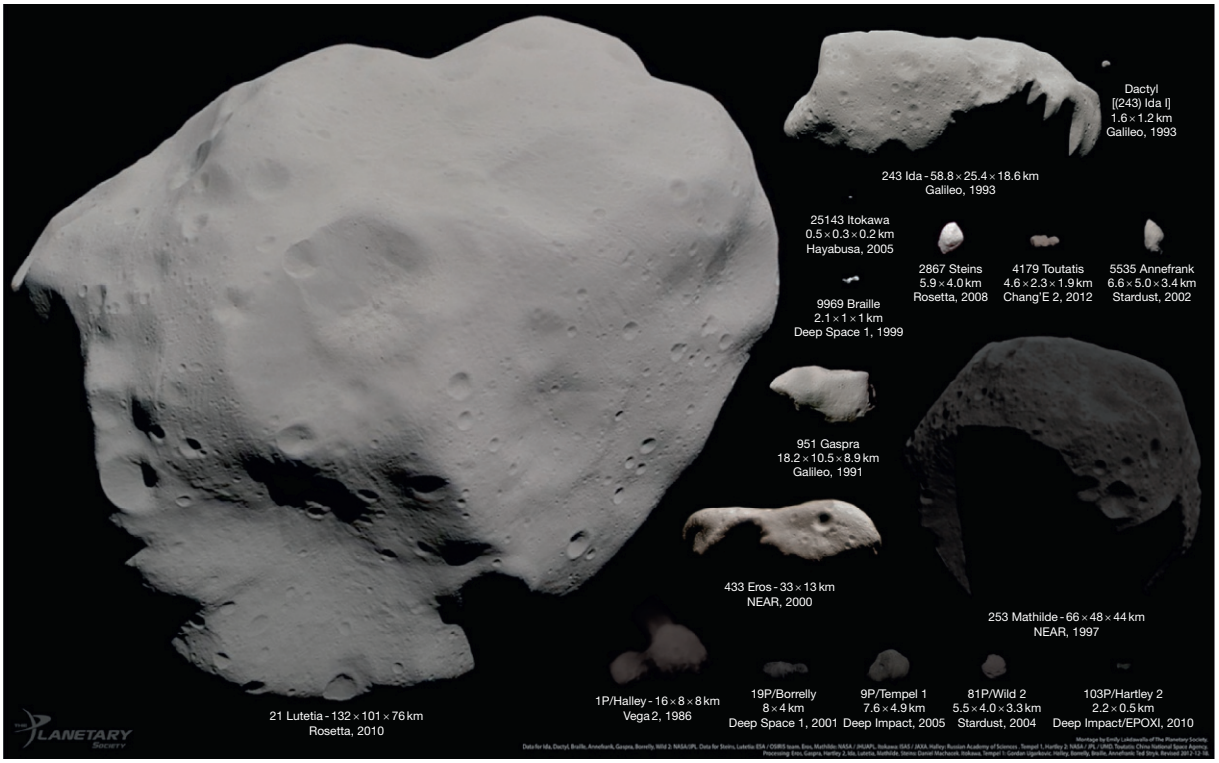


Figure 8 More complete montage of all small bodies that have had spacecraft flybys, except for Vesta, shown to scale. Each asteroid is as close to true color and albedo as possible. Note that the diameter of the largest object here, Lutetia, is about 100 km. Montage by Emily Lakdawalla. Data from NASA/JPL/JHUAPL/UMD/JAXA/ESA/OSIRIS team/Russian Academy of Sciences/China National Space Agency. Processed by Emily Lakdawalla, Daniel Machacek, Ted Stryk, Gordan Ugarkovic. Used with permission.

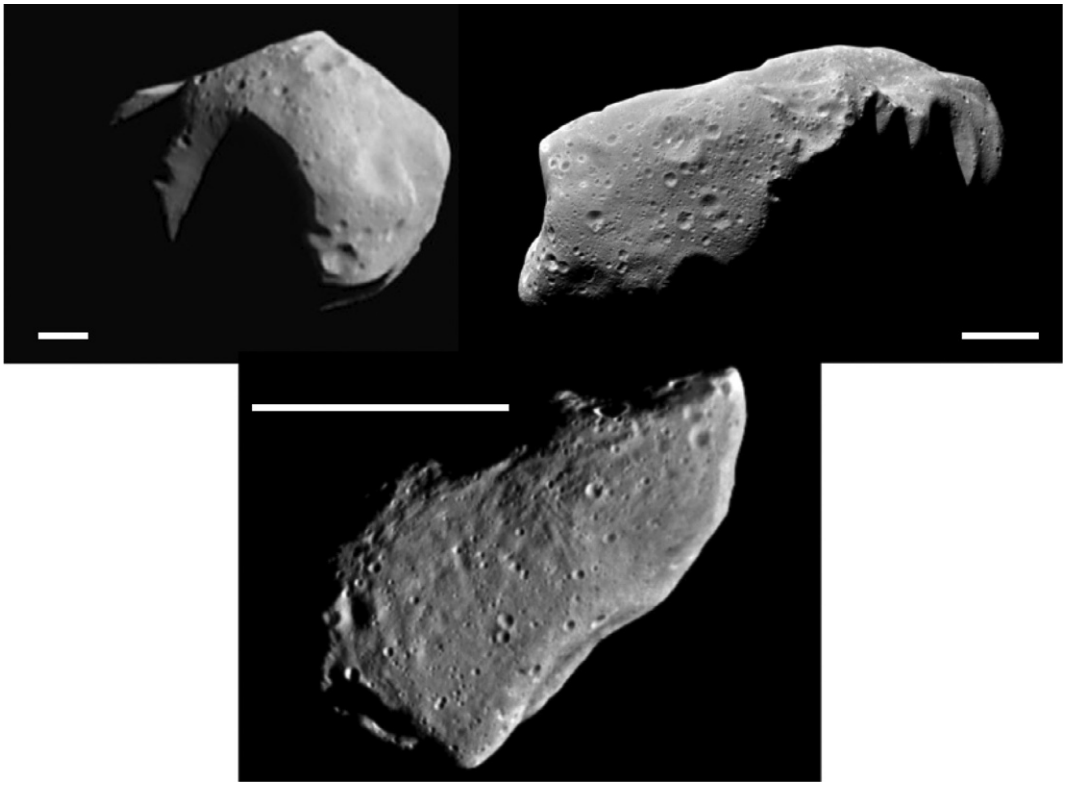


Figure 9 Higher-resolution views of some of the main-belt asteroidal flyby targets shown in **Figures 7** and **8** (clockwise from top left): Mathilde, Ida, and Gaspra. White horizontal bar in each panel measures 10 km. All pictures courtesy NASA (Photojournal images PIA02477, PIA00135, PIA00118).

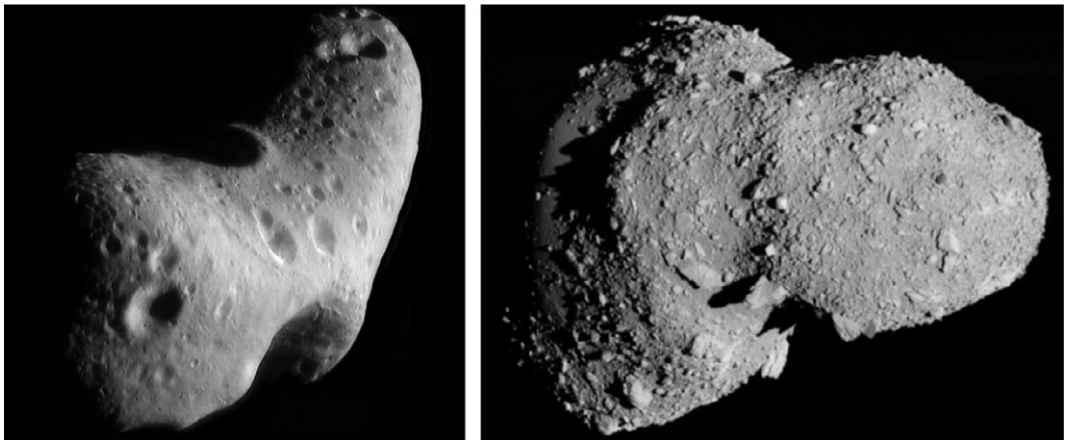


Figure 10 A comparison of Eros (left) and Itokawa (right), the two NEAs with detailed spacecraft observations. Note that Eros's long dimension is about 70 times longer than Itokawa's; Eros is a much bigger object. Also note the dramatic difference in crater density and surface texture. Eros picture credit: NASA/JHUAPL (Photojournal image PIA02923). Itokawa picture credit: ISAS/JAXA (image P-043-12079, used by permission).

with porosity. Third, the depth-to-diameter ratio (d/D) is affected by many factors including impact energy, target strength, and the degradation processes and trend (e.g., Carr et al., 1994; Melosh, 1989; Sullivan et al., 1996) and is also an indicator of regolith depth (e.g., Vincent et al., 2012a,b). The d/D on small bodies, however, is not well measured due to the uncertainties in reconstructing 3-D topography.

The cumulative SFD of craters on asteroids generally follows a power law with an index between −2 and −4, which is consistent with the SFD of main-belt asteroids from both observations (Jedicke et al., 2002) and theoretical modeling (e.g., Bottke et al., 2005; see **Figure 5**). This particular range of the power-law index is the expected result of asteroid collisional evolution (Davis et al., 2002). Starting from an asteroid population model, relative velocity, and the assumption of scaling laws, one can fit the SFD and density of craters on an asteroid or a region to derive its crater retention age, which is the time since the most recent resurfacing event (e.g., O'Brien and Greenberg, 2005; O'Brien et al., 2006). While generally this approach matches the crater size frequency on asteroids and allows us to derive these relative ages satisfactorily on, for example, (243) Ida, Dactyl (Ida's moon), parts of (21) Lutetia, and parts of Vesta (Chapman et al., 1996a; Davis et al., 1996; Marchi et al., 2012a,b), most asteroids show obvious deviations from the models at various crater sizes.

Gaspra (**Figure 9**) has few, if any, large craters but is peppered with fresh, small craters of several tens to hundreds of meters that are saturated, that is, in an equilibrium state (Carr et al., 1994; Chapman et al., 1996b). The number of craters on an object is said to reach equilibrium when for every new crater that forms, another is lost or covered up, so that the total number cannot increase. On Gaspra, the lack of large craters is inconsistent with the abundance of smaller craters. This property is so far unique to Gaspra. Gaspra possesses a few flat facets, which might be the remnants of large impact craters of sizes approaching the size of the asteroid itself (Greenberg et al., 1994; Thomas et al., 1994). The lack of large craters on Gaspra indicates a young age of ~250 My (Chapman et al., 1996b). However, Greenberg et al. (1994) and Chapman (1997) argued that the surface of Gaspra could be either much older or younger, if one assumes rubble pile structure (nearly strengthless) or monolithic structure (strong), respectively.

Unlike Gaspra, Eros (**Figure 10**) shows an extreme paucity of small craters of <100 m, while large craters >200 m are in an equilibrium state, similar to that of Ida (Chapman et al., 2002; Veverka et al., 2000). Richardson et al. (2004) and Michel et al. (2009) showed that the most likely reason for the paucity of small craters on Eros is seismic shaking erasure. The possibility that small impactors are preferentially removed by a size-dependent mechanism – that is, by the Yarkovsky effect, a thermal radiation force that causes objects to undergo semimajor axis drift due to asymmetrical radiation (Bottke et al., 2006) – has been considered but rejected (O'Brien, 2009).

Itokawa (**Figure 10**) and (2867) Šteins also show obvious depletion of small craters (Besse et al., 2012), probably also due to seismic shaking. The YORP (Yarkovsky–O'Keefe–Radzievskii–Paddack) effect (Bottke et al., 2006) – analogous to the Yarkovsky effect except it changes the asteroid's spin state due to that asymmetrical radiation of thermal emission – might be responsible for regolith movement on Šteins (Marchi et al., 2010a), effectively 'reshaping' the asteroid in that sense. Alternatively, for Itokawa, the high abundance of boulders on its surface could have protected it from forming small craters, with impactors simply fragmenting the boulders rather than creating craters (Michel et al., 2009).

The most striking features on (253) Mathilde (**Figure 9**) are several large craters with sizes comparable with the size of the asteroid itself (Cheng and Barnouin-Jha, 1999). The number of those large craters exceeds the saturation density, suggesting to several authors that Mathilde may be composed of weakly bonded, porous, crushable material that absorbs impact shock efficiently without disrupting preexisting large craters (Asphaug, 1999; Housen et al., 1999). The population of small craters on Mathilde also saturates its surface, consistent with an old surface of perhaps several billion years (Davis, 1999), although the estimate varies depending on the internal strength.

Vesta and Lutetia are both large and each shows a complicated cratering record on their surfaces. Vesta is ~560 km in diameter (Jaumann et al., 2012) and is large enough to trap

sufficient radiogenic heat during its accretion to lead to the differentiation of its interior. Vesta is thus a differentiated protoplanet (e.g., Keil, 2002; Russell et al., 2012). On the other hand, the internal structure of Lutetia (~100 km across) is uncertain, but it is unlikely to be differentiated, so it represents an ancient planetesimal (e.g., Coradini et al., 2011; Sierks et al., 2011). The cratering morphology on both asteroids includes all crater preservation states and different crater densities from region to region, suggesting very different crater retention ages from hundreds of My to >3.6 Gy (Marchi et al., 2012a,b; Massironi et al., 2012; Schenk et al., 2012; Thomas et al., 2012). On the oldest terrain on Lutetia, a region called Achaia, the SFD shows a distinct flexure point at 4–7 km diameter, attributed to the presence of a stratified surface with fractured material overlying a more competent interior (Marchi et al., 2012a), similar to the case for Mercury (Marchi et al., 2011). The full analysis of cratering properties on Vesta from Dawn data is still ongoing as of the writing of this chapter. Readers are referred to, for example, Yingst et al. (submitted for publication) and Williams et al. (in press) for the geology of Vesta.

The degradation states of craters can be gauged with the depth-to-diameter ratio (d/D), and the ratio can be also an indicator of the depth of regolith. Typically, the crater d/D on these asteroids is between 0.12 and 0.15 (Gaspra, Ida, and Eros) (Carr et al., 1994; Sullivan et al., 1996; Veverka et al., 2000), although for large asteroids Mathilde, Lutetia, and Vesta, this value could have a large range of <0.1 to >0.2 (Jaumann et al., 2012; Veverka et al., 1999; Vincent et al., 2012a,b). The small asteroid Šteins also has a large range of d/D of 0.04–0.25 (Besse et al., 2012). The smallest asteroid Itokawa has much shallower craters with d/D of 0.08 on average and a relatively narrow range of ±0.03, possibly due to the small size of the asteroid, the small number of craters with which to make the analysis in the first place, the incomplete formation of raised rims, and/or the infilling of crater by pebbles (Hirata et al., 2009; Michel et al., 2009). In contrast, the Moon, Mars, and Phobos typically have a depth-to-diameter ratio of about 0.2 (Shingareva et al., 2008). The wide range of d/D values among all these bodies indicates that crater creation and degradation depend on the detailed properties of the targets themselves. A comprehensive understanding of the variety of depth-to-diameter ratios – and of asteroid cratering in general – is a project for the future; the study of cratering on asteroids is complicated by both observational and modeling limitations. The most obvious factors include the identification of secondary craters (formed by reimpact of ejecta), the existence of extremely degraded craters underlying new craters, and the often irregular shapes of asteroids with complicated local topography. Crater chronology is almost always limited by our knowledge of the impactor populations, the mechanical properties of the surface (affecting scaling laws), and the removal process. The result is that surface ages are often strongly model-dependent.

10.15.3.2.2 Boulders

The visiting spacecraft imaged boulders on the surfaces of Eros, Lutetia, Vesta, and especially Itokawa. The surface of Itokawa is in fact dominated by boulders rather than craters (Barnouin-Jha et al., 2008; Fujiwara et al., 2006; Michikami et al., 2008). This is dramatically shown in **Figure 10**. The boulders on Vesta have not yet been thoroughly studied as of the writing of this chapter. Therefore, we will focus on the other three asteroids to discuss boulders. The shape of boulders on Eros and Itokawa is consistent with laboratory impact experiments (Michikami et al., 2010). The SFDs of boulders on all of them, as well as on Phobos, follow a power-law index of nearly −3 within a few meters to 80 m (Michikami et al., 2008; Thomas et al., 2000; 2001), a steep slope of power index of −6 for <10 cm size on Eros near its landing size (Thomas et al., 2001), and steep slopes of indexes −5 for boulders >80 m for both Eros and Lutetia (Küppers et al., 2012; Thomas et al., 2001).

The origin of boulders is generally associated with impact breaking of the target material. Since boulders are subject to destruction by the bombardment of small impactors, the existence of a large number of boulders usually suggests relatively young ages of the associated impact craters. For example, Geissler et al. (1996) noted that the distribution of large blocks on Ida is consistent with ejecta from the impact that created Azzurra crater; Thomas et al. (2001) studied the spatial distribution of boulders and their characteristics on Eros and established the formation of Shoemaker crater as the probable source of most of its boulders; a similar study for Lutetia by Küppers et al. (2012) also showed that most boulders on Lutetia were generated by the formation of the central crater in the youngest region, Baetica. The boulders on Itokawa, however, show a global distribution. There are no large craters on this small asteroid, and the total volume of boulders accounts for ~25% of the total excavated volume of craters, much higher than that on Eros (<1%) and on the Moon (~5%) (Cintala et al., 1982; Thomas et al., 2001). These arguments led Fujiwara et al. (2006) and Michikami et al. (2008) to conclude that the bulk of boulders on Itokawa most likely originated from the disruption of the parent body and re-accumulated on the surface of Itokawa during its formation.

10.15.3.2.3 Lineaments

Lineaments were first seen quite prominently on the global scale on Eros (Thomas and Prockter, 2010). The term includes grooves, troughs, and ridges, and they can be local or global in scale. Examples of lineaments are shown in **Figures 11–13**. All the asteroids that are resolved at sufficiently high resolution – except for Itokawa, which is a rubble pile – show either local or global lineaments on their surfaces. They seem to be caused by propagation of stress or tension and indicate to some extent a coherent interior. Local-scale lineaments appear to be associated with strong impacts that formed large craters. Global-scale lineaments could be associated with either the global response to seismic waves during large impacts or the stress environment of being broken up from the parent bodies (Buczkowski et al., 2008; 2012; Prockter et al., 2002; Sullivan et al., 1996; Thomas et al., 1994, 2002, 2012; Veverka et al., 1994). However, an intriguing alternate hypothesis for the Eros lineaments was given by Greenberg (2008), who suggested instead that the lineaments are near a vein of stronger rock (perhaps created by partial melt) that was more resistant to impact erosion. Grooves seen in the regolith are probably expressions of fracture within a more coherent substrate (Thomas et al., 1979) and may provide additional clues about the interior structure.

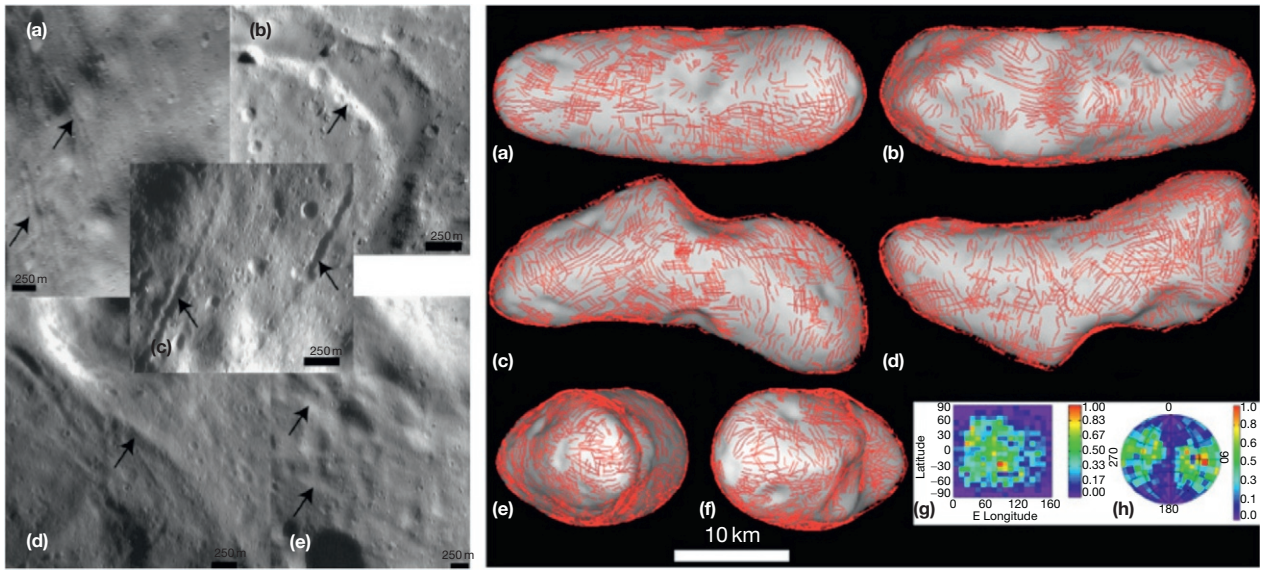


Figure 11 Lineaments on Eros. The left panel demonstrates various lineament morphologies, with the arrows indicating the relevant features, as follows: top left, grooves; top right, flat-floored troughs; bottom left, ridges; bottom right, shallow troughs; center, pit chains. The right panel shows the extensive network of 2141 lineaments (red marks) identified by software. Eros is shown in six orientations. The maps at bottom right (in two different projections) show the distribution of the poles of all the lineaments; the lineaments seem to be randomly distributed. Reprinted from Buczkowski DL, Barnouin-Jha OS, and Prockter LM (2008) 433 Eros lineaments: Global mapping and analysis. *Icarus* 193: 39–52, Figures 4 and 5, with permission from Elsevier.

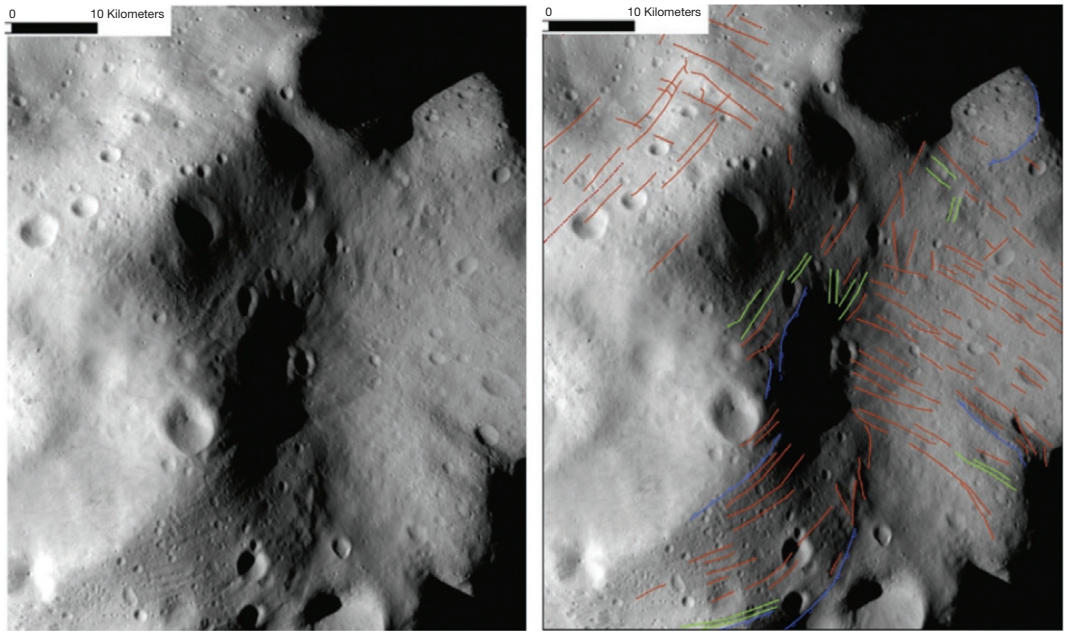


Figure 12 Lineaments on the surface of asteroid Lutetia. Note the scale bar at top. Reprinted from Massironi M, Marchi S, Pajola M, et al. (2012) Geological map and stratigraphy of asteroid 21 Lutetia. *Planetary and Space Sciences* 66: 125–136, panels (a) and (b) of Figure 6, with permission from Elsevier, with the actual image from Rosetta on the left, and the lineaments marked with colored lines on the right.

The most interesting lineament systems exist on Eros, Lutetia, and Vesta. Buczkowski et al. (2008) mapped >2000 lineaments on Eros ranging up to tens of km in length (Figure 11). They found that, except for the lineaments that are radial or circumferential to ~10 craters and obviously associated with those impacts, the majority of lineaments belong to three sets: parallel to meridian, encircling one end of Eros, and toward the other end. While interpreting the first set as due to impacts on the long side of the body, they proposed that the second and third sets could be due to the stress environment in its parent body. Nevertheless, the large number of lineaments on Eros on a global scale is not consistent with a rubble pile internal structure.

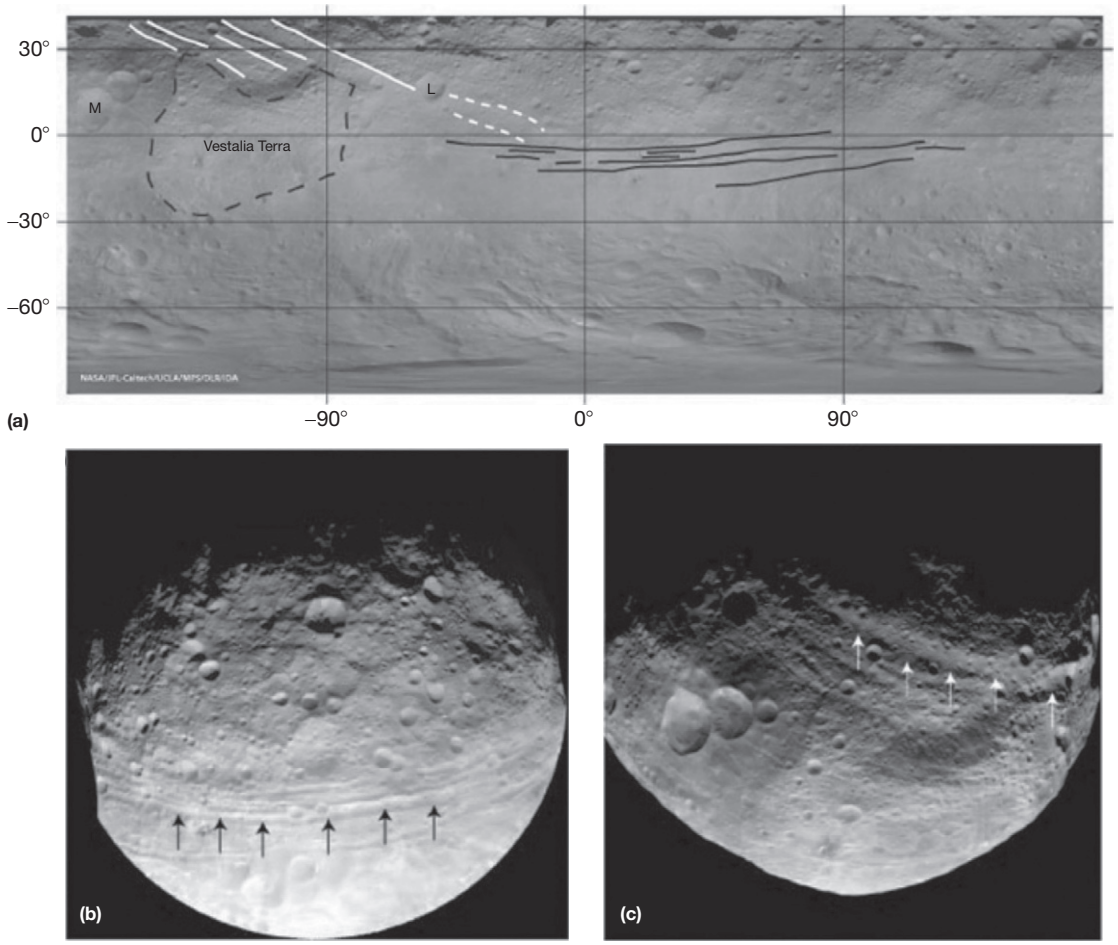


Figure 13 Lineaments on Vesta. The top panel is a cylindrical-projection map of Vesta; various systems of troughs are marked in black solid, white solid, and white dashed lines. The bottom left panel shows the equatorial troughs (the black solid lines in the top panel), while the bottom right panel shows the northern troughs (the while solid lines in the top panel). Both images are from the Dawn spacecraft. Reprinted from Buczkowski DL, Wyrick DY, Iyer KA, et al. (2012) Large-scale troughs on Vesta: A signature of planetary tectonics. *Geophysical Research Letters* 39: L18205, Figure 1, with permission from John Wiley and Sons.

Lutetia shows an extremely complicated system of lineaments on the full globe except for the geologically youngest region near the north pole (Thomas et al., 2012), and not as systematically distributed and organized as on Eros. An example of the lineament system is shown in **Figure 12**.

While the full analysis of lineament systems on Vesta is still ongoing, the most striking features are two sets of long linear structures near the equator and on the northern hemisphere that almost encircle the whole body (**Figure 13**) (Jaumann et al., 2012). Buczkowski et al. (2012) categorized these linear structures as grabens. These grabens appear to be coplanar with the respective poles coincident to the centers of two large basins near the south pole, Rheasilvia and Veneneia. They concluded that the properties of the grabens are consistent with impact-induced, global-scale deformation of a fully differentiated body, completely distinct from the responses of small, undifferentiated asteroids to impacts.

10.15.3.2.4 Smooth deposits

Both NEAs visited by spacecraft, Eros and Itokawa, display apparently smooth regions, termed 'ponds' on Eros (Abe et al., 2006; Barnouin-Jha et al., 2008; Cheng et al., 2001; Robinson et al., 2001). The smooth areas on both asteroids are located in gravitationally low areas, indicative of movements of regolith with centimeter- to millimeter-scale particles. Cheng et al. (2002) proposed that the ponded deposits on Eros are consistent with fluidlike motion of dry regolith, possibly induced by seismic agitation from impacts, the same process that is responsible for preferentially removing small craters on both of these objects (Michel et al., 2009; Richardson et al., 2004).

10.15.3.2.5 Regolith

Regolith (defined as loose, fragmental debris, regardless of origin) was not originally predicted before the Galileo encounter with Gaspra because of the low gravity (e.g., Housen et al., 1979a,b), although other early work, including that of Veverka et al. (1986), indicated that regolith would accumulate on small objects. The surfaces of asteroids are now known to possess a substantial amount of regolith derived from impacts, and this regolith demonstrates considerable mobility. The depth of regolith is generally inferred from crater morphology and ejecta, distribution of smooth deposits, the morphology of

geologic features such as boulders and lineaments, and spectrophotometric variations of the surface. In general, meters to hundred of meters thick regolith are inferred on asteroids, depending on the size of the objects (Barnouin-Jha et al., 2008; Besse et al., 2012; Carr et al., 1994; Chapman et al., 1996a,b; Prockter et al., 2002; Sullivan et al., 1996; Vincent et al., 2012a,b). The regolith on Vesta displays global variations from tens of meters to >1 km (Denevi et al., 2012; Jaumann et al., 2012) due to the complicated impact history.

Movement of regolith is evident from the localized smooth areas (as discussed in the previous section) and by mass wasting due to slope failure, as commonly seen on steep slopes of crater walls (e.g., Jaumann et al., 2012; Krohn et al., in press; Miyamoto et al., 2007; Thomas et al., 2012). The mobility and evolution of regolith are generally inferred from both geomorphological evidence and the optical maturity of regolith (i.e., from space weathering; see next section). For NEAs, the causes of mass movement are generally considered to be (a) seismic shaking induced by impact, (b) YORP spin-up, and (c) tidal disturbance from close encounters with planets. In binary asteroids, tidal effects of one body on the other can similarly move regolith around.

10.15.3.3 Space Weathering

Our knowledge about the mineralogy of asteroids is dominated by remote sensing observations of the spectral reflectance properties of asteroidal surfaces, supplemented by laboratory experiments of meteorite samples. It has long been recognized that the optical properties of silicate rocks are subject to change due to exposure in the space environment, caused by solar wind irradiation and micrometeoroid sputtering (Hapke, 1973). For the case of the Moon, the exposure causes the surface to darken, the spectral slope to redden, and the mineral absorption bands to weaken. Such optical phenomena and the associated processes are broadly termed 'space weathering.' The space weathering process on the Moon is well understood (see, e.g., Hapke, 2001; Pieters et al., 2000, for comprehensive reviews). Laboratory measurements and experiments of lunar samples showed clear evidence that vaporization of silicate minerals and redeposition of nanophase iron ($npFe^0$) coat the surface of regolith particles (cf. Hapke, 2001). Because the sizes of metallic Fe particles are smaller than visible wavelengths, they strongly absorb incident light, and their absorption coefficient decreases as the wavelength increases, resulting in darkening and reddening of the regolith particles, and masking out mineral absorption bands (cf. Hapke, 2001). However, this process cannot be easily extended to asteroids in general (Gaffey, 2010). Chapman (2004) presented an excellent historical review on this topic. Our understanding of space weathering on asteroids has been severely limited by both the lack of direct samples and the limited high spatial resolution spectroscopic data. Because mineralogical identification on asteroids almost entirely relies on the spectral reflectance properties of asteroidal surfaces, understanding the space weathering processes is an important aspect of asteroid science. As we shall see in this section, the surfaces of asteroids display different manifestations of space weathering from that on the Moon. For the convenience of our discussion, we refer to the space weathering process on the Moon and its associated optical effects as 'lunar-type.'

Space weathering on asteroids was first proposed to explain the spectral difference between the most abundant samples in our meteorite collections, the OCs, and the most abundant type of main-belt asteroids, S types (cf. **Section 10.15.1.5**). Presumably, the S types were the sources of OCs, yet the spectral differences remained problematic. Under the assumption that no mechanical or mixing process should change the mafic band centers and their relative strength, Gaffey et al. (1993) studied the mineralogy of S-type asteroids and found that the spectra of a particular subclass (known as 'S(IV)') imply mineralogy similar to the known OCs, therefore providing a generic link between S types and OCs. However, the S-type asteroids show much weaker absorption bands and much redder spectral slopes than do OCs. On the other hand, in the NEA population, a small class of asteroids, labeled the Q types, with sizes of a few km, show similar spectral characteristics to those of OCs (Binzel et al., 2002; McFadden et al., 1984). This relationship is a good reality check because the OC meteorites were NEAs before falling to Earth, so one would expect that there would be a closer spectral match. Binzel et al. (1996, 2002) showed that nonprimitive NEAs show a continuous spread in their spectra from OC-like to S-like, implying a process that is continuously operating on the surfaces of one class and changing their optical properties to the other. Both observational (Binzel et al., 2010) and dynamic (Nesvorný et al., 2010) evidence suggested that tidal encounters to within ~5 planet radii of Earth and Venus could reset the weathered surfaces on small S-type asteroids, by exposing a relatively fresh unweathered surface and changing them to Q types.

To put it another way, S-type main-belt asteroids have highly weathered surfaces because they have been exposed to the space environment for ostensibly tens of millions to billions of years. Q-type surfaces, that is, those with little or no weathering, on the other hand would appear in the NEA population because NEAs are collisional fragments and relatively recent escapees from the asteroid belt; thus, their surfaces have only been exposed for millions of years or perhaps much less.

Recent observations of small S-type asteroids in the main belt that are of comparable size with NEAs also show spectral variation, which is logical because these objects have fresher surfaces and may be on their way to becoming NEAs. Some of these small objects have spectra that match the OCs (Mothé-Diniz et al., 2010). Further studies of the colors of many S-type dynamic families in the main belt have shown that their red spectral slopes are clearly correlated with dynamic ages since breaking up, consistent with continuous weathering process on S-type asteroids (Nesvorný et al., 2005).

Recent asteroid missions have both confirmed the existence of space weathering on asteroids and raised more questions. The Galileo spacecraft flybys of two S-type main-belt asteroids Gaspra and Ida showed that the surfaces of Gaspra and Ida both have large variations in colors and at least a factor of two variation in the strength of mafic bands. While the overall spectra of Gaspra and Ida are typical of S-type asteroids, small, fresh craters and ejecta appear to be bluer, and the ridges with steep slopes tend to be bluer than other areas (Belton et al., 1992, 1994; Geissler et al., 1996; Sullivan et al., 1996).

Close examination of the spectral trend on Ida revealed that the geomorphologically freshest units have spectral intermediate between OCs and the S-type asteroids (Chapman, 1996). However, because the albedo variation on Ida is small, the space weathering process does not seem to have darkened the surface of Ida as much as it has of the Moon.

The clearest evidence of space weathering on Eros was identified in the large crater named Psyche, where brighter materials are exposed on the wall and darker materials are deposited at the bottom (Clark et al., 2001). The albedo variation from wall to bottom is a factor of several. The color of the bright crater wall is slightly bluer than that of the dark crater floor, consistent with, but much subtler than, lunar-type space weathering (Izenberg et al., 2003; Murchie et al., 2002). Therefore, the wall of Psyche crater was recently exposed due to mass wasting of the original regolith on the slope and is much less space-weathered than the crater floor. However, there are a number of inconsistencies in terms of space weathering on Eros. First, the color and spectral variations on Eros are remarkably uniform to within 2% (Izenberg et al., 2003; McFadden et al., 2001). The lack of color contrast between the supposedly fresh crater wall and the weathered crater floor certainly suggests different weathering processes on Eros than on Gaspra, Ida, and the Moon; this is represented schematically in **Figure 14** (Gaffey, 2010). Second, the downslope movement of Psyche crater wall material had to be driven by, for example, other impacts after the formation of the slope. However, Psyche crater appears to be the youngest on Eros; none of the small craters on Eros appear fresh, that is, brighter and bluer than average surface. Cheng et al. (2002) considered that the shaking of Eros, probably induced by tidal forces during close approach with terrestrial planets, could induce mass wasting on slopes and the formation of ponds. In addition, ejecta blocks on Eros, mostly excavated by the impact that formed Shoemaker crater, show similar albedo and spectra to other areas on Eros, a relation that also indicates relatively mature surfaces.

The Hayabusa mission to Itokawa provided evidence of space weathering on this small (550 m) NEA from both spectroscopic observations and the first samples directly returned from known sites on an S-type asteroid. Hiroi et al. (2006) showed that a dark region on Itokawa has a spectrum clearly consistent with significant space weathering with accumulated $npFe^0$ (nanophase metallic iron) with about 0.03–0.07 vol.%. Ishiguro et al. (2007) analyzed the spectral data of the whole surface of the asteroid to map the space weathering maturity on Itokawa and found that fresh materials are generally observed in regions of steep slopes and craters, whereas mature materials are ubiquitously distributed. Laboratory measurements of the Itokawa samples showed surface modifications on the rims of some particles, with sulfur-bearing Fe-rich nanoparticles, interpreted as a consequence of alteration by vapor deposition (Noguchi et al., 2011, 2014).

The presence of $npFe^0$ is one consequence of space weathering that appears to be the same on both the Moon and asteroids. In addition to the evidence of $npFe^0$ found in Itokawa regolith samples (Noguchi et al., 2011, 2014), nanosize metallic particles are also found in some OC meteorites (Longobardo et al., 2011; Moretti et al., 2005, 2006) and in the howardite (i.e., likely Vesta fragment) Kapoeta meteorite (Noble et al., 2010). Furthermore, extensive laboratory experiments of space weathering have been conducted to simulate the effects of micrometeoroid impacts through laser impulses and solar wind bombardment through plasma irradiation (e.g., Brunetto and Strazzulla, 2005; Brunetto et al., 2006; 2007; Dukes et al., 1999; Hapke, 1973, 2001; Loeffler et al., 2008, 2009; Marchi et al., 2005; Noble et al., 2007; Strazzulla et al., 2005; Yamada et al., 1999). The most significant result from these experiments has been the reproduction of the lunar space weathering effects by generating $npFe^0$ coating on mineral grains (Keller and McKay, 1993). The timescale of space weathering due to micrometeoroid impacts has been estimated to be of order 10^8 years (Sasaki et al., 2001).

For non-S-type asteroids, such as the more primitive objects in the C-complex, the effect of space weathering is not obvious. This lack of spectral detection might be because the dark, opaque, carbonaceous materials (which are thought to be the dominant composition of low-albedo asteroids like those in the C-complex) may obscure the effects of space weathering. Or alternatively, a low abundance of Fe might suppress the formation of $npFe^0$ and slow down the process of lunar-type space weathering.

Finally, we point out the puzzling case of Vesta, which resides on the opposite end of the traditional space weathering problem, that is, Vesta shows a lack of weathering, even though it is expected. McCord et al. (1970) found an exact match between the spectrum of Vesta and the spectra of the most abundant achondrite meteorite samples, the howardite–eucrite–diogenite (HED), leading to the conclusion that Vesta is the source of HEDs. Compelling evidence from almost every aspect confirms the HED–Vesta connection. Early evidence includes the identification of small (few km) spectrally similar

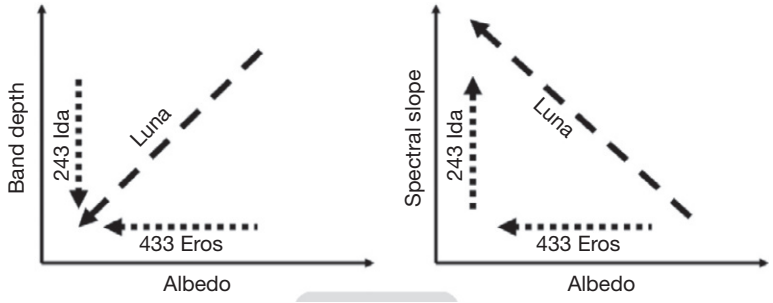


Figure 14 Comparison of space weathering across asteroids and the Moon. There are many manifestations of 'space weathering,' and the trends of albedo versus depth of an absorption band versus spectral slope can be different depending on the properties of the surface under study. Reprinted from Gaffey MJ (2010) Space weathering and the interpretation of asteroid reflectance spectra. *Icarus* 209: 564–574, Figure 4, with permission from Elsevier.

asteroids (now known as taxonomic type V) that provide a dynamic pathway to transport meteorites from the asteroid belt to Earth (Binzel and Xu, 1993) and the existence of a large basin near Vesta's south pole (Thomas et al., 1997a,b). Recent observations by the Dawn mission confirmed the mineralogical connection between Vesta, HEDs, and V-type asteroids (Ammannito et al., 2013; De Sanctis et al., 2012, 2013) and confirmed a highly impacted surface with two large, overlapping basins near the south pole, with the youngest one perhaps just ~1 Gy old (Marchi et al., 2012b; Schenk et al., 2012). However, laboratory studies suggest that ion irradiation on HEDs darkens and reddens the samples rapidly, corresponding to a space weathering timescale of 10^5–10^6 years (Fulvio et al., 2012). The lack of space weathering on Vesta itself is thus a mystery (see summary by Pieters et al., 2006). To explain it, Vernazza et al. (2006) postulated that Vesta has a magnetic field that can shield it from solar wind irradiation. The discovery of magnetism in the Millbillillie eucrite (Fu and Weiss, 2011) seems to support this idea. But magnetic shielding does not completely protect the whole surface of Vesta from weathering, and Dawn should have identified local weathered areas with high-resolution imaging and spectroscopy. Alternatively, the whole surface of Vesta could have been recently resurfaced by a major impact. However, the age of the largest impact basin on Vesta is much older than the timescale generally accepted for space weathering, as mentioned in the preceding text. On the other hand, the comparisons of red slopes among V-type asteroids and HEDs support the existence of space weathering on V-type asteroids, although the reddening trend with solar wind exposure age is difficult to interpret (Marchi et al., 2010b). Pending the solution of whether the traditional view of space weathering on the Moon and S-type asteroids occurs on Vesta, Pieters et al. (2012) proposed an alternative view of the space weathering based on Dawn observations, in which a physical process on the surface such as regolith mixing, rather than chemical alteration, dominates the space weathering. Still, the most significant question for Vesta is that we do not know how to fit the lack of traditional space weathering into the whole paradigm of such processes on the Moon, Eros, and Ida.

10.15.3.4 Surface Morphology on Cometary Nuclei

The encounters with comet 1P/Halley by ESA's Giotto spacecraft (Reinhard, 1986) and the Soviet Union's Vega spacecraft (Sagdeev et al., 1986) marked the first close-up imaging of solar system small bodies. During the close encounter, the Halley Multicolor Camera (HMC) on board the Giotto spacecraft centered on the brightest part of the inner coma, showing the silhouette of a large, solid, and irregularly shaped nucleus and jet-like dust activity that was much brighter than the nucleus (Keller et al., 1986). Due to the bright jets in the foreground of the nucleus, it was difficult to see the surface morphology of this cometary nucleus from these images. Nevertheless, the solid nature of cometary nuclei as icy conglomerates as proposed by Whipple (1950) was directly proved correct. And for the first time, the albedo of a cometary nucleus was directly measured to be ~4%.

Since the P/Halley flybys, there have been five additional cometary missions returning disk-resolved images of four more cometary nuclei (**Table 1**). A montage (not to scale) of the comets is shown in **Figure 15**. The imaging data from those cometary missions have formed the foundation of our current understanding of the surface morphology of cometary nuclei. While the comets have a variety of overall morphologies, there are many common features, including pits and pitted terrains, smooth areas, icy patches, and numerous small, bright, or dark spots.

Comets have numerous (nearly) circular depressions with or without raised rims on their surfaces, generally termed 'pits' due to their completely different morphology from impact craters on asteroidal surfaces. Pits are observed on all four cometary nuclei with various sizes and morphologies (**Figure 15**; Britt et al., 2004; Brownlee et al., 2004; Thomas et al., 2007, 2013a,b). The associated pitted terrains usually occupy a large fraction of a comet's surface, except in some regions where there are smooth terrains. The cumulative size distributions of pits on comets 9P/Tempel 1 and 81P/Wild 2 have slopes between −1.7 and −2.1 (Thomas et al., 2013a), significantly different from the size distribution of typical impact craters on the Moon or asteroids (between −2 and −4). Therefore, pits on comets either do not all originate from impacts or have been modified after formation. The pits on comet P/Tempel 1 are up to a few hundred meters in diameter and up to 25 m in depth, mostly without flat floors. Those on P/Wild 2 have larger sizes up to ~1.5 km, mostly with flat floors and some with central peaks. The most distinctive feature on comet P/Wild 2 is the nearly vertical walls of the pits, sometimes with overhangs. While some pits with raised rims on some comets could have impact origins (such as the two circular depressions bracketing the Deep Impact (DI) mission impact site on comet P/Tempel 1), most of them are probably cryovolcanic collapse features associated with cometary activity (Belton and Melosh, 2009; Belton et al., 2008). Belton et al. (2013) ascribed >90% of the pits to cometary outburst activity.

The apparently smooth areas on comets P/Tempel 1 (Thomas et al., 2007, 2013a) and 103P/Hartley 2 (Thomas et al., 2013b) are probably one of the most intriguing features observed; such areas are visible in **Figure 15**. These terrains have smooth texture up to 30 m scale but most likely also at a 5 m scale, with slightly lower reflectance than the surrounding terrains, and they are usually laterally confined. They occupy topographic lows with slopes up to a few degrees, suggestive of flows. The 'mesas' on comet 19P/Borrelly as termed by Britt et al. (2004) appear to be similar to the smooth areas on comets P/Tempel 1 and P/Hartley 2. Belton and Melosh (2009) proposed a fluidized multiphase transportation of dust resulting from the sublimation of material with higher volatility than water, such as CO or CO_2, from beneath the surface as the origin of the smooth areas on comet P/Tempel 1. No such smooth area was observed on comet P/Wild 2. Belton (2010) used the smooth areas to hypothesize an evolutionary sequence of the four comets, where the surface of comet P/Wild 2 represents the early stage in the sequence and has not developed smooth areas that are large enough to be visible, while the surface of comet P/Hartley 2 represents the latest stage with the largest fraction of surface covered by smooth areas.

From the Deep Impact flyby spacecraft images of comet P/Tempel 1, Sunshine et al. (2006) for the first time unambiguously discovered water ice deposits on the surface of a cometary

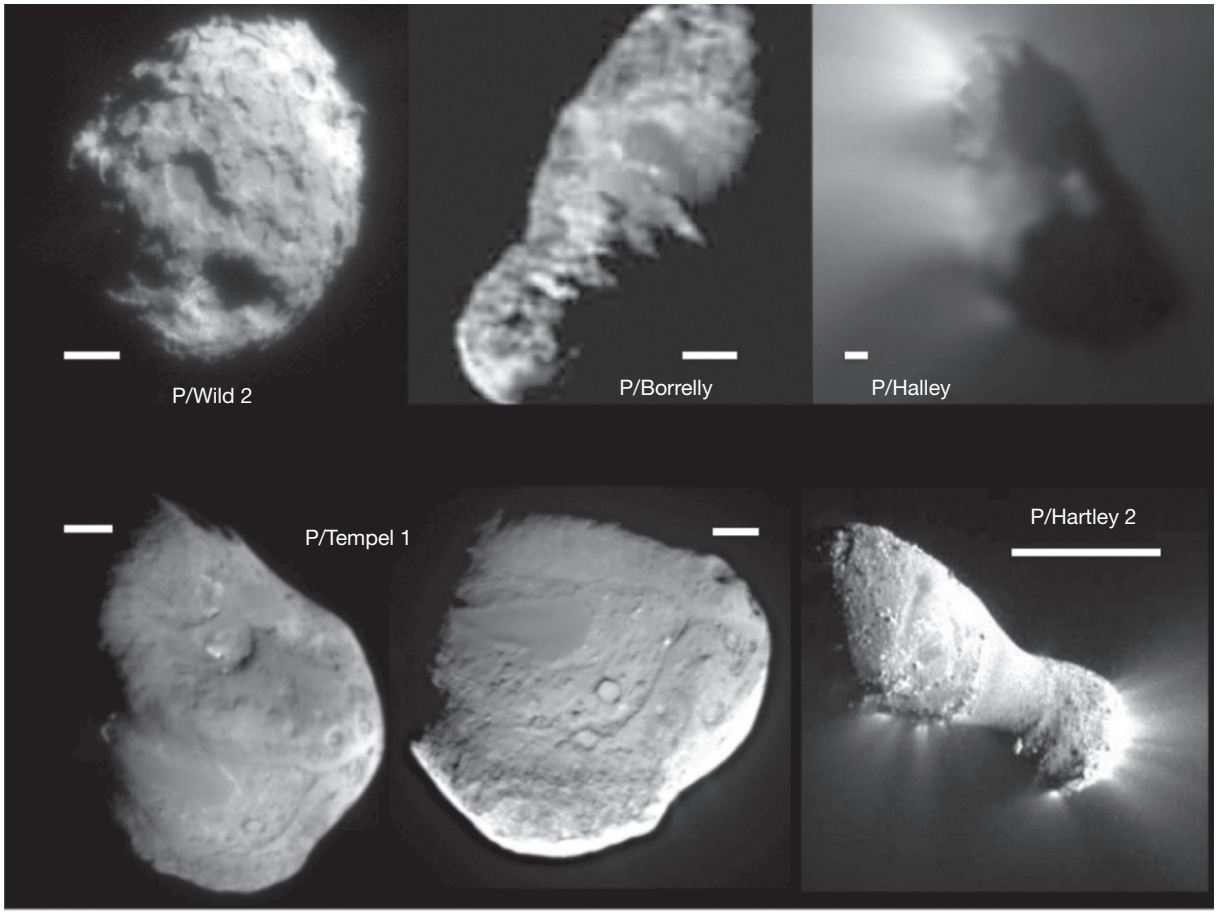


Figure 15 Montage of the five comets that have had spacecraft encounters (with imaging). P/Tempel 1 has been visited twice. P/Halley was visited four times (see **Table 1**) but for clarity we only show a Giotto picture. White horizontal bar in each panel represents 1 km. Note the variety of surface morphologies. Image sources: P/Wild 2 is courtesy NASA/JPL-Caltech (Photojournal image PIA05571); P/Borrelly is courtesy NASA/JPL (Photojournal image PIA03500); P/Tempel 1 (left image) is courtesy NASA/JPL/UMD (Photojournal PIA02142); P/Tempel 1 (right image) is courtesy NASA/JPL-Caltech/Cornell (Photojournal image PIA13860); P/Hartley 2 is courtesy NASA/JPL-Caltech/UMD (Photojournal image PIA13570); P/Halley is from the Giotto archive of data in the NASA Planetary Data System (Keller et al., 1992).

nucleus. Similar water ice concentration was again observed on comet P/Hartley 2 (Sunshine et al., 2011). The icy patches cover less than 1% of the total surface areas of comets and only contain about 3–6% water ice, with typical particle sizes of ~30 μm, much larger than those in ejecta observed by Deep Impact (~1 μm) (Sunshine et al., 2007) and in comet P/Hartley 2's coma (Protopapa et al., 2014). Temperature measurements and thermal modeling suggest that the water ice deposits are thermally decoupled from the refractory dust on the surface (Groussin et al., 2007, 2013). The locations of the identified ice deposits on comets P/Tempel 1 and P/Hartley 2 are both near the morning terminators, suggesting that they are probably condensation of water vapor in diurnal cycles as opposed to ice exposed from the interior (which would have much smaller particle sizes, too; Sunshine et al., 2007). It is quite possible that the reason for water ice being definitively discovered on only two cometary nuclei is that only the Deep Impact spacecraft, among those having visited comets, is equipped with a spectrometer capable of detecting water ice. Therefore, it is reasonable to speculate that water ice patches are likely common on cometary nuclei.

In addition to the large-scale features discussed earlier, there are numerous bright and dark spots on all well-imaged cometary nuclei (**Figure 16**). Some of them appear to be albedo features, such as some bright spots on comets P/Wild 2 and P/Tempel 1. The spots could be water ice concentrated in areas due to topography, but it is not possible to determine their origins conclusively because of the limited resolution or the unavailability of spectroscopic data. On the other hand, some dark spots could be either small-scale albedo features or small, deep pits or holes (Nelson et al., 2004).

Cometary surfaces are under constant change due to outgassing activity. The second flyby of comet P/Tempel 1 by the NExT mission in 2011 just one perihelion passage after the Deep Impact flyby in 2005 showed clear change in surface morphology (**Figure 17**; Thomas et al., 2013a), although no obvious change in photometry was identified (Li et al., 2013). The most significant change observed was the retreat of the scarp bounding the smooth area near the south pole by at least 50 m. In addition, at least two crudely triangular areas evident in 2005 disappeared by 2011, representing a backwasting along at least 1000 m of the boundary. It is estimated that

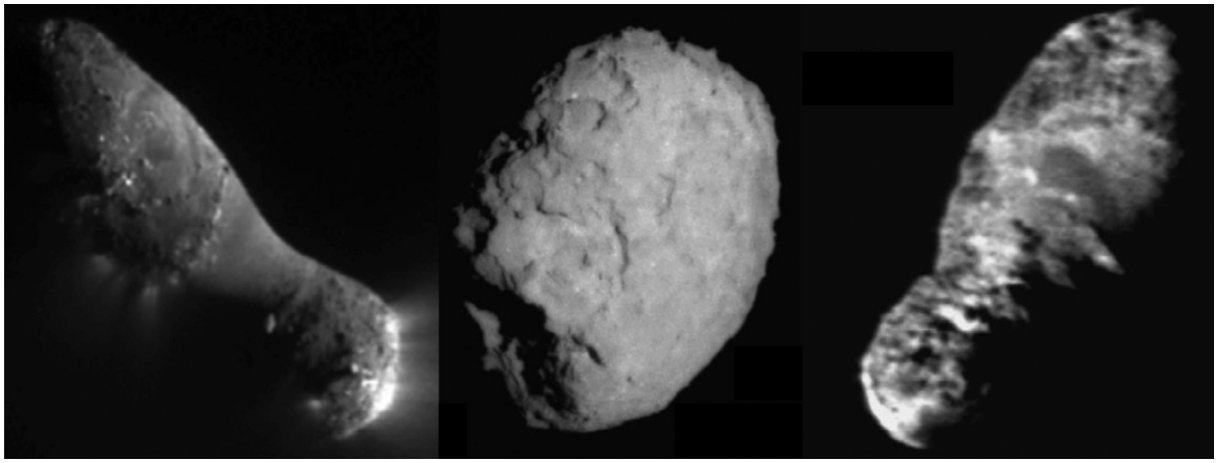


Figure 16 Examples of the bright and dark spots seen on cometary surfaces. Images are of P/Hartley 2, P/Wild 2, and P/Borrelly, from left to right. The P/Hartley 2 and P/Borrelly images are the same as those seen in **Figure 15**, but with better contrast to enhance the features. P/Wild 2 image is from the Stardust archive of data in the NASA Planetary Data System (Newburn and Farnham, 2008). Some of the bright spots could be footpoints to jets. The dark spots could be low-albedo features or actual pits.

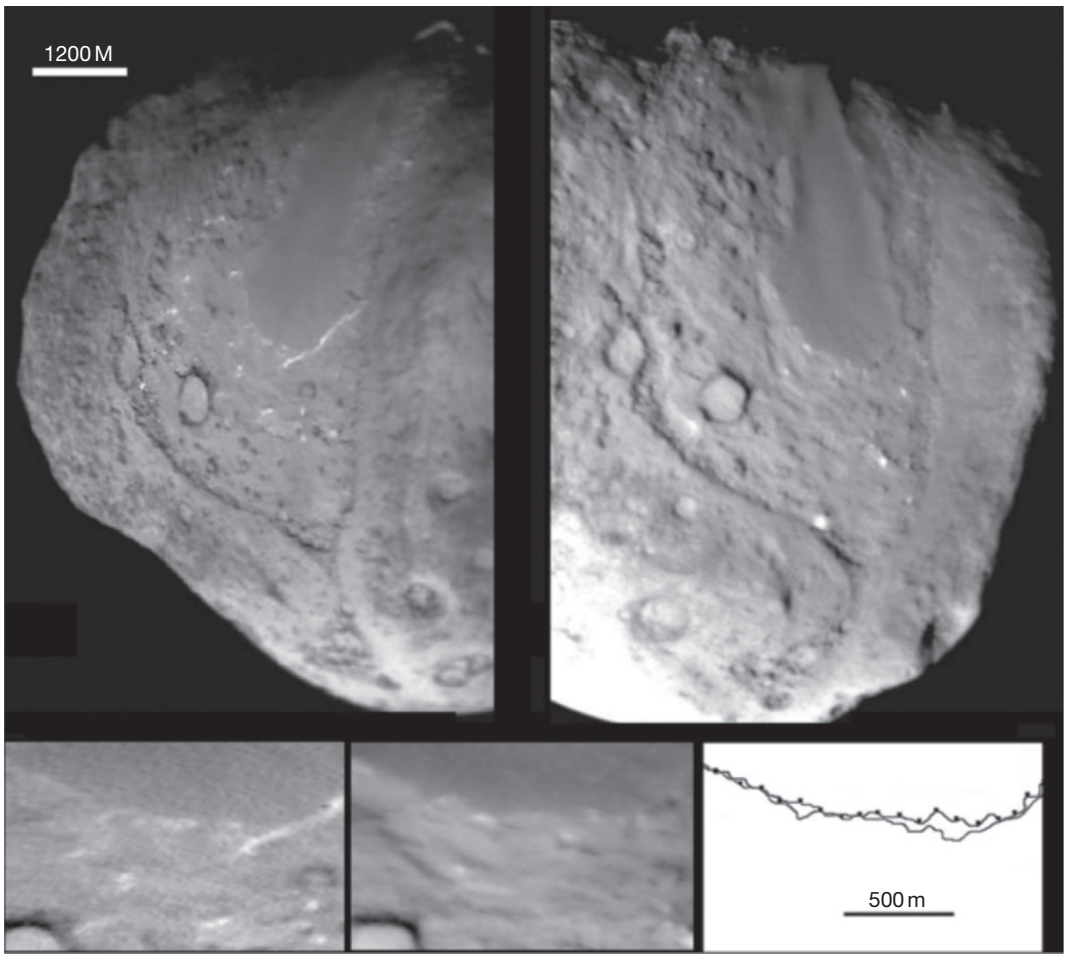


Figure 17 Changes in the surface of comet P/Tempel 1 from one orbit to the next. The left image is from the Deep Impact flyby in 2005, and the right image is from the Stardust flyby in 2011. Various bright and dark spots have appeared and/or disappeared in the intervening 6 years. The edge of the scarp has clearly changed, losing up to 50 m of extent in some locations. This change is shown in the bottom three panels, with the right panel showing the traces of the scarp. Reprinted from Thomas P, A'Hearn M, Belton MJS, et al. (2013) The nucleus of Comet 9P/Tempel 1: Shape and geology from two flybys. *Icarus* 222: 453–466, Figure 9, with permission from Elsevier.

the total volume loss was about $2–4 \times 10^5$ m^3, corresponding to $8–16 \times 10^7$ kg, assuming an average density of 0.47 kg/ m^{-3} (Richardson et al., 2007). It has also been noted that there is concentrated jet activity originating near the scarps (Farnham et al., 2013). In addition to the boundary of the smooth features, several small bright albedo spots (<30 m) in the region have changed in contrast and extent. However, due to their small sizes, the changing viewing geometry, and the different imaging instruments, it remains uncertain whether these are real changes or are due to the effect of different illumination and viewing geometries or instrument effects. Interestingly, no obvious ejecta blanket produced by the DI impact could be identified on the surface of comet P/Tempel 1 (Schultz et al., 2013).

The high-resolution images of comets P/Borrelly, P/Wild 2, and P/Tempel 1 reveal another common feature, that is, ubiquitous surface layering that possibly extends into the interior. Based on these layers, Belton et al. (2007) hypothesized a 'talps' model or 'layered pile' model to describe the internal structure of JFCs. In this model, the nucleus interiors are composed of layers of different thicknesses, sizes, and possibly compositions that were accumulated during the primordial accretion phase of comets through low-speed collisions between cometesimals. This hypothesis presents a completely different internal structure from the classic rubble pile model for asteroids (cf. Richardson, 2002, and references therein). While the evolution of asteroids mostly occurred in the inner solar system and was dominated by intensive collisions after their formation, leading to their rubble pile nature, the evolution of comets probably took a completely different path. JFCs we see today were frozen for ~4 Gy in the Kuiper Belt, that is, where they are thought to have formed, before being gravitationally perturbed into the inner solar system (e.g., Duncan and Levison, 1997; Duncan et al., 2004; Morbidelli and Brown, 2004). Comets in JFC orbits are active for only 7% of the orbit (Duncan et al., 2004). Therefore, the evolution of JFCs is first dominated by the collisional environment in the Kuiper Belt and then on their surfaces by volatile sublimation during repeated perihelion passages. If talps are primordial in JFCs, then their preservation seems to indicate a much more benign collisional history than the asteroids have undergone. Future cometary missions, especially the CONSERT experiment on ESA's Rosetta mission that will use ground-penetrating radar to study the internal structure of comet 67P/Churyumov–Gerasimenko (Schulz, 2009), should provide definitive tests on the talps model.

10.15.4 Structure, Interior, and Spin

10.15.4.1 Structure of Asteroids

From the discovery of the largest asteroid, (1) Ceres, in 1801 by Piazzi (Foderà Serio et al., 2002) until the 1990s, asteroids were largely unresolved point sources. Reflected light can reveal surface composition and perhaps some level of structure through the thermal properties, but the interiors of asteroids were largely inaccessible to study. As described in the previous section, spacecraft missions can reveal exquisite detail of a body's surface structure and in some cases interior by means of gravity measurements. However, the number of missions will by necessity be limited, so studies of the population must rely largely on ground-based remote sensing.

Ground-based radar does reveal the shapes and surface details of NEAs at few-meter resolution and has led to some surprising discoveries about their structure. Ostro (1989) summarized the early efforts to study the shapes and outer contours of NEAs. The two planetary radar facilities, Arecibo Observatory and Goldstone Deep Space Network, are complementary in their capabilities and ability to study solid solar system bodies, primarily NEAs (Ostro et al., 2002 and references therein). Starting in 1998, the increased sensitivity of the Arecibo planetary radar system, along with the huge increase in the discovery rate of NEAs, has begun to reveal their detailed shapes. Some were the expected collisional fragments, with irregular facets, craters, and ridges. However, some were surprisingly spheroidal, even at sizes of only a few hundred meters. The self-gravity of such small bodies is not nearly enough to overcome the strength of the rock, so these must be strengthless rubble piles. However, other objects in the same size range were seen to be elongated, or rapidly rotating, which required a coherent rock. The diversity of NEAs is remarkable and indicates a variety of formation mechanisms (Nolan et al., 2005; Ostro et al., 2002).

Radar imaging is accomplished by sending a monochromatic signal to the asteroid and receiving the signal, which reflects from the solid surface. Because the precise timing and characteristics of the transmitted signal sent are known, the return signal reveals much about the body, depending only upon knowing the time and speed of light. We can use the time sampling and rotational velocity information to reconstruct an image of the asteroid, with some projection effects, and ambiguity in that several points on the asteroid surface may have identical distance and apparent velocity and thus map to the same location in the radar image (see Magri et al., 2007; Ostro, 1993, for more details).

The resolution of a radar image depends on the signal-to-noise ratio (SNR) of the signal reflected from the asteroid. The power of the transmitter, size of the telescope, and gain contribute to the SNR, but the dominant factor is usually distance from Earth. The SNR depends on the inverse fourth power of the distance, so NEAs dominate the sample of high-resolution images. About 300 objects have been imaged as of 2014, with about 50 of those at the highest resolution available at Arecibo, 7.5 m per pixel. Before 2011, the highest resolution at Goldstone was 19 m per pixel, but currently, for very close approaches to Earth, images from GSSR can reach 3.75 m per pixel as in asteroid 2005 YU_{55}. Examples of some radar images are shown in **Figure 18**. As described earlier, NEAs are collisional fragments of larger objects that are then perturbed into planet-crossing orbits (e.g., Bottke et al., 2002a). While some NEAs do indeed look like irregular collisional shards, many others do not. About 25% of the NEAs between 200 and 800 m are spheroidal, and another 10–15% are two spheroidal lobes joined together. Elongated objects, with axial ratios of 2:1 or more, comprise about 10% of the sample, while binary systems are another 15% of NEAs less than 10 km in diameter (Taylor et al., 2012). An asteroid less than 10 km can only be spherical from its own gravity if it is strengthless, although surface features such as craters and ridges can be supported by friction in the surface material. Surface features seen in radar images must

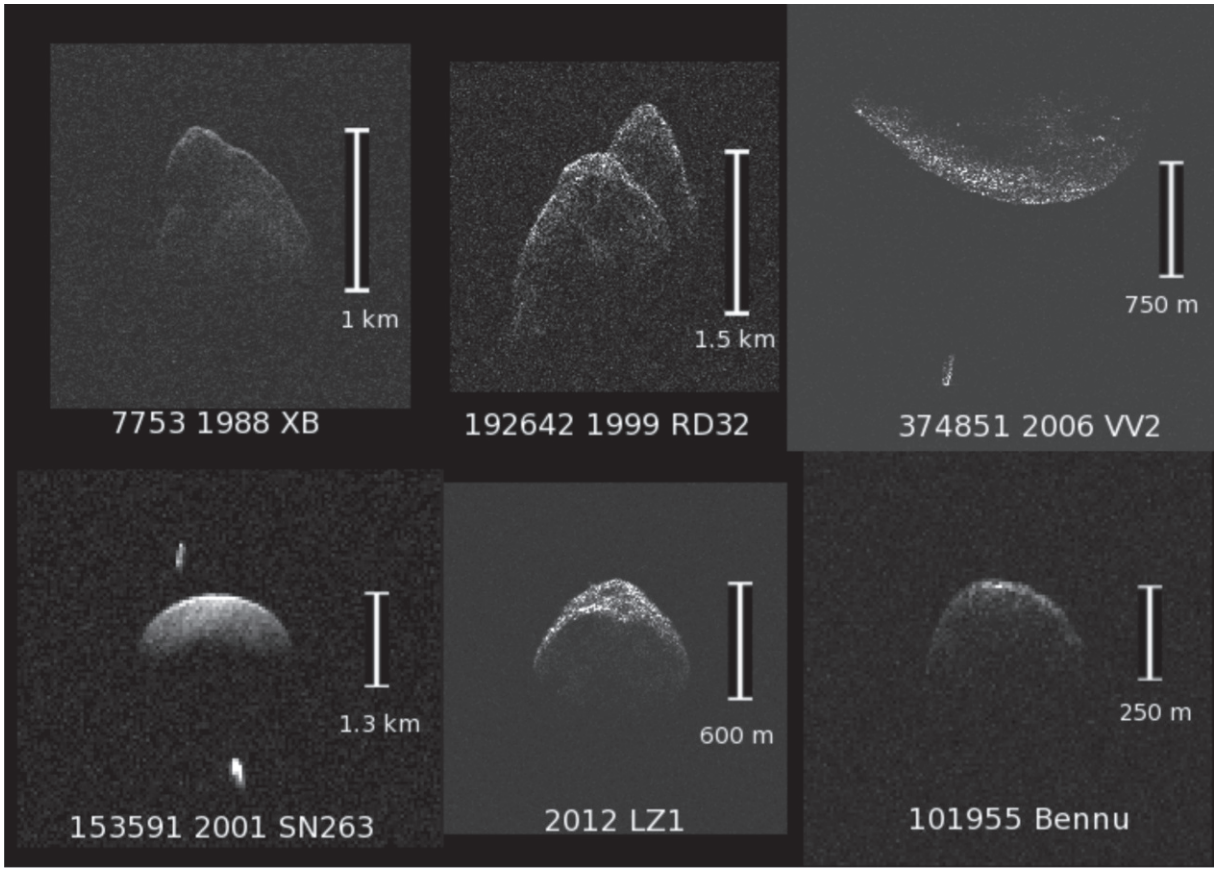


Figure 18 Arecibo radar images of a variety of asteroids can reveal surface features as small as 7.5 m. Scale bars show the approximate size in the vertical direction. In each of these images, the distance from Earth increases from top to bottom (except for 2006 VV2, where it is the opposite), and Doppler frequency or object rotation velocity in the line of sight increases from left to right. These are not spatial images, but for a rigid body, they map into a spatial image, and surface features can be interpreted readily with some projection effects.

be interpreted with caution, because the image is not a spatial image, but it does map into a spatial image. A radar image is really distance from the observer in one dimension, and rotational velocity in the other dimension, or distance from the rotation axis. By convention, radar images always appear as if the object is viewed from the north pole, illuminated from 90° phase angle. From a single image, the location of a feature on the surface cannot be determined uniquely. However, a series of images over several days can be used to derive a shape model (e.g., Magri et al., 2011; Nolan et al., 2012; Ostro et al., 2005). More limited datasets can be used to identify surface textures and features that often reveal internal structure.

The sample of objects shown in **Figure 18** includes some notable objects. Asteroid (153951) 2001 SN_{263} is the first triple asteroid system discovered among the NEAs, where the outer satellite is larger (below the primary in this image) orbiting in about 6.2 days and the smaller inner one (above the primary) orbits in about 16.5 h. The OSIRIS-REx mission will visit asteroid (101955) Bennu (formerly 1999 RQ_{36}), a 550 m spheroidal asteroid, which although rotating relatively quickly (4.2 h) does not appear to have a satellite larger than 10 m at the present time. Both (374851) 2006 VV_2 and Bennu have one or more boulders on the surface larger than 10 m, perhaps similar to those seen on Itokawa and Eros. The population of NEAs is varied in size and shape, composition, and rotation, suggesting a variety of possible outcomes for the same formation conditions.

10.15.4.2 Spin of Asteroids

The rotation rate of an object can put constraints on its internal structure. Optical light-curve observations of asteroids and comets have been carried out by both amateur and professional astronomers for many decades and provide an extensive database with which to study the population of asteroids. The most reliable data from the asteroid light-curve database compiled by Warner et al. (2009) – and which is frequently updated at http://www.minorplanet.info – are plotted in **Figure 19** for main-belt asteroids, with H magnitude on the x-axis and rotation frequency (rotations per day) on the y-axis. The H magnitude is the brightness of an asteroid at 1 AU from both Earth and Sun, and at zero phase angle, so is a proxy for the asteroid diameter. The plot shows that for about $H < 20$, a rotation frequency of 10, or about 2.4 h rotation period, is a limiting value for all asteroids. This corresponds roughly to the fastest rate at which a strengthless rubble pile structure can spin without flying apart (e.g., Pravec and Harris, 2000). Smaller asteroids ($H > 20$) can break this spin barrier and can rotate as

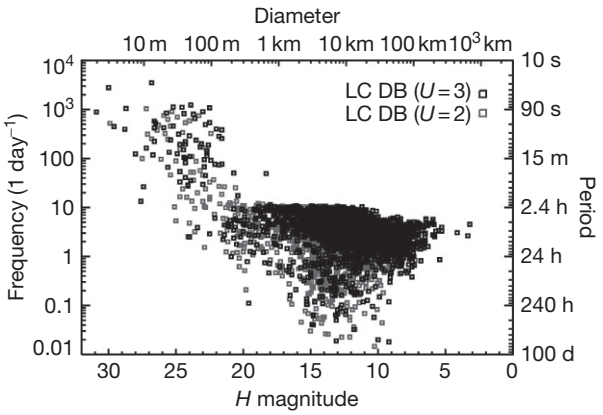


Figure 19 Distribution of the rotation rates of asteroids. Rotation rates are given as frequencies (left axis) and periods (right axis). The rates are shown as a function of diameter (top axis), which is estimated from the absolute magnitude (*H*; bottom axis) using an assumed geometric albedo of 0.20 (cf. **Figure 5**). Plotted here are the higher-quality observations ('*U*' parameter of 2 or 3) taken from the asteroid light-curve database (Warner et al., 2009), based on many nights of observation for each asteroid to determine the rotation rate from changes in the brightness of the object as it rotates. Note the sharp cut in the distribution at a rotation period near 2.4 h. This rate seems to be a limit for asteroids larger than *H* magnitude of about 21, which has been interpreted as being the fastest that a strengthless, rubble pile body can spin without breaking apart. Figure is courtesy Patrick Taylor.

rapidly as 30 s. This certainly requires some tensile strength, although some internal fractures and porosity can still be present. The spin limit depends on the material density, so it is slightly slower for icy bodies. No asteroids appear to be both small and rotating slowly, although these are difficult to measure using ground-based optical telescopes. Radar measurements should see small objects rotating slowly if they exist, but the distribution from radar observations matches the light-curve distribution very closely, despite having very different observational biases.

10.15.4.3 Binaries

The discovery of the first binary NEA, (185851) 2000 DP_{107}, in 2000 (Margot et al., 2002) opened up a new window on studying interior structure of asteroids. Radar images allow the sizes of both components to be observed so the mass of the primary body can be derived from the orbit. Searches for binary asteroids had been inconclusive prior to this time, even leading to theoretical arguments as to why binaries did not exist and would not be stable (e.g., Gehrels et al., 1987). Light-curve measurements suggested binary systems existed for decades, but the possibility of two-lobed or contact binary objects could not be ruled out for most of the cases. Clear evidence of occultations and eclipses in asteroid light curves had not been confirmed prior to the discovery of 2000 DP_{107}. However, after the first radar detection of an NEA binary system, showing the clear separation of the components, the light-curve evidence was then considered much more convincing.

As of 2014, two-thirds of the approximately 50 known NEA binary systems have been discovered using radar, still the most powerful means of identifying binary systems. Light-curve detection continues to be an important contributor and is now showing binary systems are equally common in the main-belt population in the few kilometer-size range (Pravec et al., 2006). Yet another technique that works quite well for binary discovery in the main belt is direct imaging through the use of telescopes that have adaptive optics. Normally, the atmosphere distorts the wave fronts from an astronomical source that are collected by a telescope; turbulence in the atmosphere leads to the light coming through many cells of air that each have slightly different indices of refraction. When the telescope tries to bring all the light to a focus, the resulting image is blurry, usually far blurrier than would be expected from just considering the diffraction limit of the telescope's optics. For example, a 10 m telescope at a good astronomical site, without the aid of adaptive optics, might normally achieve images with ~0.5 arcsecond resolution, whereas the diffraction limit (Rayleigh criterion) at visible wavelengths for such a telescope is about 0.01 arcsecond. 'Adaptive optics' is the term for the technological feat of having a deformable mirror in the optical path (after the secondary mirror) that deblurs the light by changing the light's reflection just enough to cancel out the effects of the turbulent cells of air. The mirror is deformed and re-deformed many times per second. An example of the incredibly good resolution that can be achieved through adaptive optics is shown in **Figure 20**, where one can clearly see a triple asteroid system, (216) Kleopatra (Descamps et al., 2011).

Following the discovery of binary 2000 DP_{107}, several formation theories were put forth. Tidal effects due to planetary encounters, collisional debris, disruption by collisions and reformation as a binary, and rotational fission all seemed viable possibilities (Merline et al., 2002). However, after several more binary systems were observed, and their common features compared, it was clear that some of these formation theories could be ruled out. All NEA binary systems have a fast rotating primary body, usually close to 2.5 h, near the theoretical spin barrier for a rubble pile structure (see **Figure 19**). The angular momentum of the system was well above the stability threshold, suggesting that spin-up and fission had occurred. The size of the secondary is remarkably consistent, between one-third and one-fifth of the primary diameter, and the orbit falls within 3–5 primary radii away. Many of the satellites are in synchronous rotation, and most are in very circular orbits, suggesting substantial tidal evolution. However, a few systems are not in synchronous rotation, and some satellites have orbital eccentricities of up to 0.15; one system, (153958) 2002 AM_{31}, has eccentricity near 0.45 (Taylor et al., 2013).

Ostro et al. (2006) and Scheeres et al. (2006) discussed the shape and orbital dynamics of the binary asteroid, (66391) 1999 KW_4, in exquisite detail from radar observations. The elongated shape of the satellite demonstrated libration, and material shifting on the surface due to the extremely low gravity explains the smooth central band of the primary. Many other asteroids have now been shown to have this same basic shape, not all of them binary systems, which suggests either that a satellite may have formed and was subsequently lost or that this shape can be the result of spin-up alone and does not require the presence of a satellite to form initially. Future spacecraft missions to such objects may help answer this question.

The diversity of binaries among asteroid taxonomic types suggests that mechanical structure, and not composition, is the

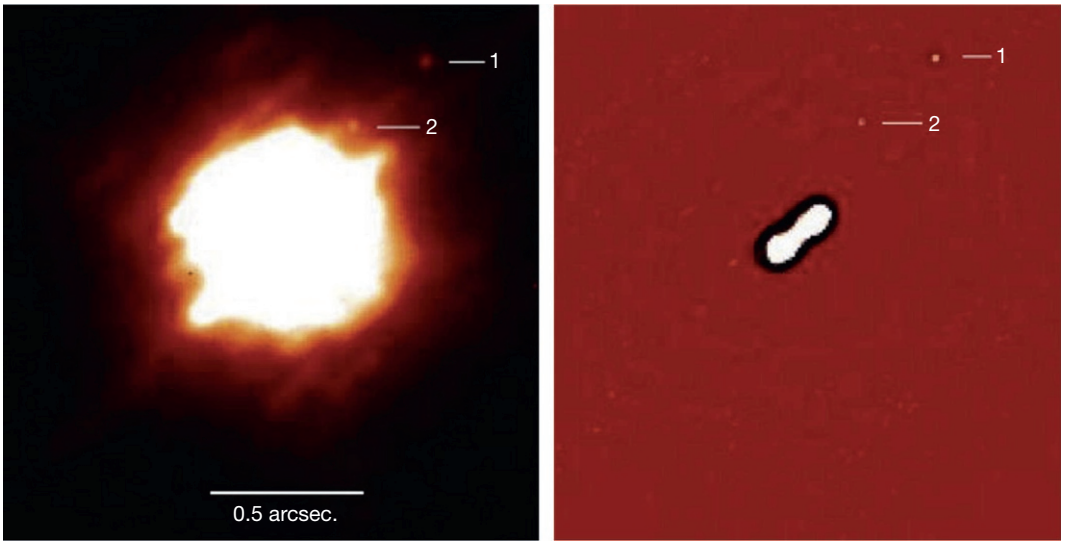


Figure 20 Adaptive optics image of asteroid Kleopatra as taken by Descamps et al. (2011) with the Keck II telescope atop Mauna Kea, Hawaii. The panel on the left shows the original image, which has been heavily stretched to show the effects of diffraction. The panel on the right shows the image after some deconvolution and filtering and has a resolution of about 0.035 arcsecond. The elongated nature of the asteroid is evident, and Kleopatra's two moons are also clearly revealed. Note that the effective resolution in the right panel rivals the resolution obtained by, for example, the Hubble Space Telescope. Reprinted from Descamps P, Marchis F, Berthier J, et al. (2011) Triplicity and physical characteristics of Asteroid (216) Kleopatra. *Icarus* 211: 1022–1033, Figure 1, with permission from Elsevier.

most important factor in forming binaries. Binary systems are found, for example, among S-complex objects, X-complex objects, and V types (e.g., 2006 VV_2). In particular, some binaries share characteristics that point to formation by spin-up due to the YORP effect, in particular some V types and Xe types, especially those Xe types that are part of the Hungaria family (the high inclination, low semimajor axis, and high-albedo group seen in the upper left of **Figure 3**). Note that Xe types used to be known as E types in the Tholen system and may have enstatite surfaces.

The densities of NEA binaries, 1.0–2.0 g cm^{-3}, imply porosity of 30–50% if the grain density is 2.5–3.2 g cm^{-3} for ordinary or carbonaceous meteorites assumed to be the parent body analogs for at least some of the S-complex and X-complex objects. Britt and Consolmagno (2003) measured porosities of meteorites and found that cracks and fissures on all scales give moderate porosities (20–30%) for seemingly solid samples. Asteroid porosities of 50% or even more are not inconsistent with many meteorite samples we have. A thorough compilation of asteroid porosities is given by Carry (2012) and shown in **Figure 21**.

The first asteroid shape model resulting from radar observations, that of (4769) Castalia (formerly 1989 PB), is a two-lobed structure, reminiscent of a molar tooth (Hudson and Ostro, 1994). Many more asteroids have since been seen to have this same basic shape, not only NEAs but also comet nuclei (Harmon et al., 2010, 2011; Magri et al., 2011). Unlike near-Earth binary systems, these contact binaries tend to be slowly rotating and have nearly equal mass lobes (Benner et al., 2008). Jacobson and Scheeres (2011) suggested that when an object splits into nearly equal mass lobes, a stable binary system is difficult or impossible to form. The equal mass systems, (69230) Hermes (Margot et al., 2003) and (90) Antiope, are rare in the near-Earth and main-belt zones. In the Kuiper Belt, binary systems of nearly equal mass are more common, but the discovery biases are quite different from those in the other regions, as are the size ranges observable. The conditions that would lead to the formation of a contact binary rather than a rapidly rotating binary with mass ratio 1:3 or more may not really be fundamentally different, but may instead be a stochastic process.

10.15.4.4 Cometary Properties

Comets lag far behind asteroids in terms of our understanding of ensemble properties of structure and rotation. There are fewer available targets for study by radar, and rotational studies at visible wavelengths can be complicated by the fact that often, the coma is much brighter than the nucleus. Radar delay-Doppler imaging of the sort that appears in **Figure 18** has only been obtained on four comets: P/2005 JQ_5 (Catalina) (Harmon et al., 2006), 73P/Schwassmann–Wachmann 3 (Howell et al., 2007), 8P/Tuttle (Harmon et al., 2010), and P/Hartley 2 (Harmon et al., 2011). This last comet is shown in **Figure 22**, and one can compare the radar image of P/Hartley 2 to those shown in **Figures 15** and **16**.

In addition to the delay-Doppler imaging, basic radar spectra have been obtained from 11 other comets total (Harmon et al., 2004, 2011), yielding constraints on nuclear sizes and on rotation rates. In some cases, the radar echoes contain a wide (in velocity space) skirt of flux. Since the wavelength of the radar is cm scale, this skirt is thought to show echoes from very large grains (>2 cm diameter) in the near-nucleus coma. Typical cometary activity can only lift (through the drag force) grains up to cm size; larger grains are usually too heavy and do not reach escape velocity (Grün and Jessberger, 1990), although in the context of fragmentation, larger grains can be liberated (e.g., Howell et al., 2007). In any case, the radar echo

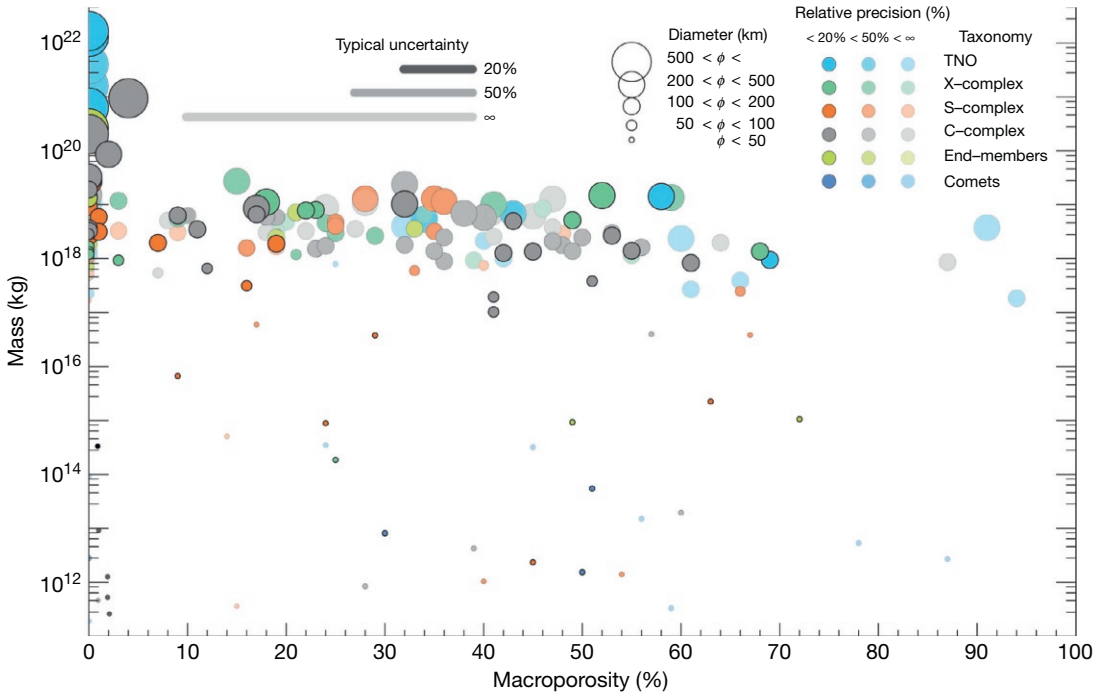


Figure 21 Scatter plot of macroporosities and masses of various small bodies as compiled by Carry (2012). Shading of each circle indicates the approximate uncertainty of the porosity measurement; Carry (2012) quantized the uncertainty as 20%, 50%, and 'infinity' (i.e., very large). Most of these results are from the 1990s and later; it is only since then that we have come to realize that many small bodies have large porosity. Reprinted from Carry B (2012) Density of asteroids. *Planetary and Space Sciences* 73: 98–118, Figure 8, with permission from Elsevier.

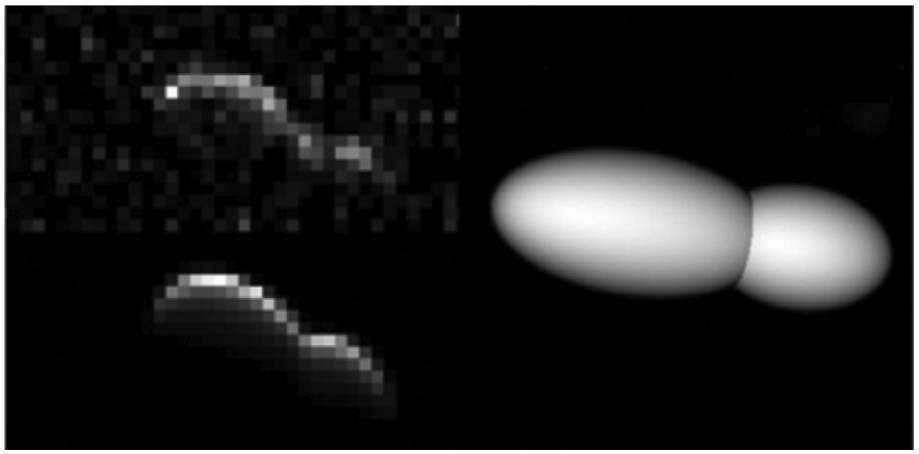


Figure 22 Radar echoes from and shape modeling of comet P/Hartley 2. One of the delay-Doppler images of the nucleus is shown at top left, along with a simulated image from a shape model of two joined ellipsoids at bottom left. That shape model itself for the nucleus is shown at right. The long axis of the nucleus is about 2.3 km. Compare the model to the spacecraft imaging shown in **Figures 15** and **16**. Radar observations prior to the flyby of the comet by the Deep Impact spacecraft helped to refine the position of the comet nucleus and were critical to the success of the mission. The radar image sequence also helped to refine the rotation state. This is an excellent example of how ground-based observations provide important context for spacecraft missions. Reproduced from Harmon JK, Nolan MC, Howell ES, Giorgini JD, and Taylor PA (2011) Radar observations of comet 103P/Hartley 2. *Astrophysical Journal Letters* 734: L2, Figure 2(b), ©AAS, with permission.

skirt is showing us the large end of the comet's dust coma and can provide an important constraint on the total mass loss.

A review of cometary rotation is provided by Samarasinha et al. (2004), and from 2004 to 2014, there have been only a few new rotation periods found for comets. The existence of excited (nonprincipal axis) rotation states has been noted in a few comets (e.g., comet P/Halley; Belton et al., 1991), but it is not always easy to confirm such spin states. One important advance since the mid-2000s however has been the discovery that several comets change their rotation periods significantly on orbital timescales. This phenomenon was first confirmed for 10P/Tempel 2 (Knight et al., 2012; Mueller and Ferrin,

1996), although it might have been seen in other comets earlier (e.g., Schleicher et al., 1991). Recently, changes in the rotation period have been noticed in comets 2P/Encke, P/Tempel 1, and P/Hartley 2 (Chesley et al., 2013; Meech et al., 2011; Mueller et al., 2008; Samarasinha et al., 2011). The change is on the order of 0.1–1%. For P/Encke; the change is a few minutes per orbit; for P/Tempel 1 and P/Hartley 2, it is on the order of 1 h per orbit. The primary driver is a net torque exerted on the nucleus by the reaction force of its outgassing. The physical importance of understanding such short-term changes in the rotational state is that one can then try to understand how long it takes for the excited rotation state to damp out, an effect that depends on the bulk solidity of the nucleus itself – that is, how much different pieces of the low-density nucleus rub up against each other. In other words, studies of cometary rotation states provide a clever way to try to understand the physical nature of cometary interiors.

The rapid changes that a comet can make to its own rotation period mean that in general, a comet has no memory of its primordial spin state. An intriguing study – but likely only feasible far in the future – would be to look at the distribution of rotation states among LPCs that are each on their first trip in from the Oort cloud and before they actually turn on. In other words, how are the LPCs spinning before they feel any torques from their own outgassing? Such observations would likely give us a window into a more primordial spin state distribution or at least certainly a glimpse of what very long-term, evolutionary dynamic processes might affect the comets while they are in the Oort cloud, tens of thousands of AU from the Sun. The difficulty is that there are virtually no LPCs discovered that are new in the Oort sense that are inactive; such objects are simply too faint to detect with current technology.

Even such fundamental quantities as the mass and the density of a cometary nucleus have been exceedingly difficult to constrain. Spacecraft have in general been too far from each comet under study to have their trajectories gravitationally deflected. The ejecta cloud produced by the impactor experiment with the Deep Impact mission suggested that the density of comet P/Tempel 1 is very low, about 400 kg m^{-3} (Richardson et al., 2007) although the quoted range is 200–1000 kg m^{-3}. The shape of comet P/Hartley 2's nucleus, like two ellipsoidal lobes, suggests that it might be showing an equipotential surface; if that is the case, Thomas et al. (2013b) derived a density of 200–400 kg m^{-3}. Before these missions, the most common way to estimate a comet's mass was through the use of its nongravitational parameters, that is, by making use of the changes to a comet's orbit caused by the reaction force of its own outgassing. Not only does outgassing change the spin state (as discussed earlier), but also the orbit then deviates from a purely gravitational solution (Marsden et al., 1973). Some recent estimates have been made for P/Wild 2 (upper limit of about 600–800 kg m^{-3}; Davidsson and Gutiérrez, 2006) and P/Borrelly (100–300 kg m^{-3}; Davidsson and Gutiérrez, 2004). Other estimates using the fact that the density must be sufficient to keep a comet from splitting apart due to its rotation are in the range of about 200–500 kg m^{-3} (Davidsson, 2001). While it is wise to keep in mind that some of these analyses are model-dependent, the results are consistent with the idea that cometary nuclei are underdense and very porous bodies.

10.15.5 Composition

10.15.5.1 Basic Premises

It is commonly assumed that because the comets we see today have spent most of their lifetimes in an environment too cold for most geochemical processes to occur and are too small to differentiate or retain heat of formation, that to first order the gas that comes off pristine comets will tell us about the compositional and chemical environment of the young solar system's protoplanetary disk. The situation is more complex for asteroids, which may have suffered more processing due to internal heat and in some cases differentiation, but some primordial information can still be inferred.

As discussed earlier, the comets we see today are vestiges of the icy planetesimals that were formed during the era of planetary formation. When such a planetesimal was created, while the primary component was water, many icy species (CO, CO_2, CH_3OH, CH_2O, etc.) were incorporated into its body. The relative amounts of these species were a direct sample of the chemical and thermal properties of the protoplanetary disk at that time and at that location (or perhaps the locations sampled by the planetesimal if scattering occurred while it was accreting).

Since the volatile content of asteroids in general does not outgas, but is locked in the minerals, it is much more difficult to ascertain compared with comets. Except in the case of sample return, asteroid composition is determined either by analysis of meteorites (which may have lost some of their volatiles during atmospheric entry or due to weathering after reaching the ground) or by modeling the compositional diagnostics in the reflectance and emission spectra of the surfaces. These multicomponent fits are highly dependent on the typical grain size of the regolith on the surface, making it difficult to distinguish between, for example, adsorbed water in layer phyllosilicates, OH groups on hydrated silicates, and actual water on clay grains. Basically, most remote sensing observations – even by a spacecraft that might be considered to be in situ – are only sensitive to the topmost microns of an object's surface, leaving most of the bulk of the object itself unsampled. Radio and radar observations can penetrate deeper – several wavelengths, so roughly a few decimeters – but there are fairly restrictive detectability limits in that wavelength regime.

As mentioned in Section 10.15.1, some asteroids have been observed to have outbursts of dust, showing something like a cometary coma and tail. While some of these objects have merely suffered a collision and thrown off dust as debris from an impulsive impact, several objects have been observed to have sustained dust production over weeks or months. Presumably, such activity is driven by water (although this has yet to have observational confirmation), and rather than the sharp classifications of icy comets and rocky asteroids, it seems likely there is a spectrum of compositions between the two endpoints. The active objects with asteroid-like orbits have been dubbed the 'main-belt comets.' The prototype MBC, with not only the comet designation 133P/Elst–Pizarro but also the asteroid designation (7968), is shown in Figure 23. Dynamic modeling of the long tail shows that the cause of the ejection of dust from the object could not have been impulsive; it must have lasted for months and therefore is likely to be water sublimation-driven. Furthermore, P/Elst–Pizarro has shown activity consistently now at

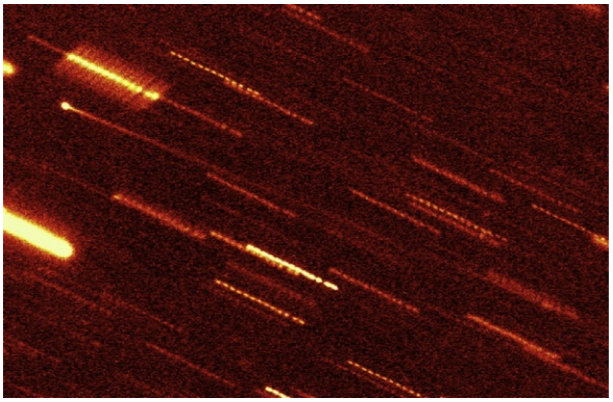


Figure 23 Image of comet 133P/Elst–Pizarro, a.k.a. Asteroid (7968) or the first 'main-belt comet,' as observed by Hsieh et al. (2004). The comet is the continuous object with the bright spot (the comet's head) in the upper left and the long tail extending down toward lower right. Other dotted sequences of spots in the image are background stars; note that the comet's tail is in a different orientation from the star trails. The comet's tail extends several arcminutes and is continuous, giving strong evidence that the object was continuously active for months on end and thus being driven by cometary activity. While about ten objects in the asteroid belt have shown extended emission from dust, about seven (including P/Elst–Pizarro) show evidence for comet-like activity; the others were likely the result of impulsive events such as from an impact event. The persistence of ice to drive such activity has significant implications for our understanding of small-body evolution. Reproduced from Hsieh HH, Jewitt DC, and Fernández YR (2004) The strange case of 133P/Elst-Pizarro: A comet among the asteroids. *Astronomical Journal* 127: 2997–3017, Figure 4, ©AAS, with permission.

nearly the same orbital longitudes for four orbits (Hsieh et al., 2013). There are currently less than ten known MBCs, but surveys are ongoing and it is expected more will be discovered. Statistical analysis of asteroid populations has suggested there may be on the order of a few hundred to a few thousand MBCs (Bertini, 2011; Gilbert and Wiegert, 2010; Hsieh, 2009). The existence of these objects has already transformed the way we think about the formation of planetesimals and mixing in the early solar system (see, e.g., Bertini, 2011). As we move from the discovery phase to the determination of composition of these objects, they should prove to be a valuable tool in constraining protoplanetary chemistry.

10.15.5.2 Volatiles

As discussed in Section 10.15.1, the primary difference between comets and asteroids is that the composition of comets is dominated by volatiles, while asteroids are dominated by rock-forming minerals but should have volatiles in increasing proportion with distance from the Sun. As comets approach the Sun, their volatiles begin to sublimate, likely from the outer 1–2 m (Prialnik et al., 2004). This produces a coma that is usually bright enough to make it difficult to discern the nucleus itself from Earth-based observations. Thus, frequently, the nucleus's properties are inferred from studies of the coma. From observations of gas in cometary comae, we can examine the distribution, composition, and rate of gas production from the nucleus. Modeling of these properties can yield, for example, the rotational state of the nucleus and lower limits on its size (e.g., Samarasinha et al., 2011; Schleicher and Woodney, 2003). Some gases sublimate directly from the nucleus (parent) and others are produced in the coma (daughter), either by chemical reactions or, more commonly, by photodissociation of the original molecules. As water is the most abundant volatile, all other volatiles in comets are typically measured by abundance relative to water or its direct photofragment OH.

The various volatile species found in comets have a wide variety of sublimation temperatures, which means they have the ability to begin outgassing at a variety of heliocentric distances. CO is the most volatile with a sublimation temperature of 25 K, though the somewhat more abundant CO_2 with its sublimation temperature of 80 K and ability to begin outgassing around 13 AU is believed to be the likely driver of most distant comet activity (Meech and Svoren, 2004). Water gradually takes over as the primary driver of cometary activity as heliocentric distance decreases. This variation in onset sublimation temperature leads to variations in abundance with heliocentric distance for some species (**Figure 24**; Biver et al., 1997, 2002). This is not the only reason abundances can vary with heliocentric distance, for instance, the variation in the HNC/HCN ratio has been shown to be due to chemistry in the coma (Irvine et al., 1998; Rodgers and Charnley, 1998).

Chemical abundances of volatiles vary from comet to comet, but when a large number of comets are observed, statistically significant groupings appear; thus, much work from the 1980s to the 2010s has gone into developing and understanding compositionally based taxonomies. The first of these was done with daughter fragments OH, NH, CN, C_3, and C_2 as these are readily observed in the optical, often with smaller aperture telescopes (~1–2 m), owing to their electronic transitions at visible wavelengths. A'Hearn et al. (1995) observed 85 comets, discovering that a significant fraction of JFCs (those originating in the Kuiper Belt) were depleted in carbon-chain molecules – C_2 and C_3 – relative to comets originating in the Oort cloud. This is shown in **Figure 25**, which is taken from the work by A'Hearn et al. As of 2013, this work has been extended to 101 comets, and principal component analysis continues to show that the carbon-chain depletion is not associated with evolution, but is primordial in nature (D. Schleicher, private communication). Additional optical surveys have been done spectroscopically. Those done by Fink (2009) and Langland-Shula and Smith (2011) also divide their comets into taxonomies and agree with the A'Hearn et al. results where they overlap on carbon-chain depletion.

To determine composition, one would ideally observe the species that sublimate directly from the surface – that is, the parent species – rather than the chemical fragments. Most of these species, however, only have spectroscopic transitions in the infrared (vibrational and rovibrational) and radio (rotational). Some important species – most significantly, CO_2 – have transitions that occur at wavelengths where the atmosphere is effectively opaque. Such species can only be observed from space. Emissions from many parent species at radio wavelengths tend to be faint enough that they can only be observed in the brightest comets. Despite these challenges, the database of observed comets at these wavelengths continues to grow and taxonomies of parent species are beginning to emerge. **Figure 26** from the work of Mumma and Charnley (2011) shows all of the species thus far observed. Radio surveys

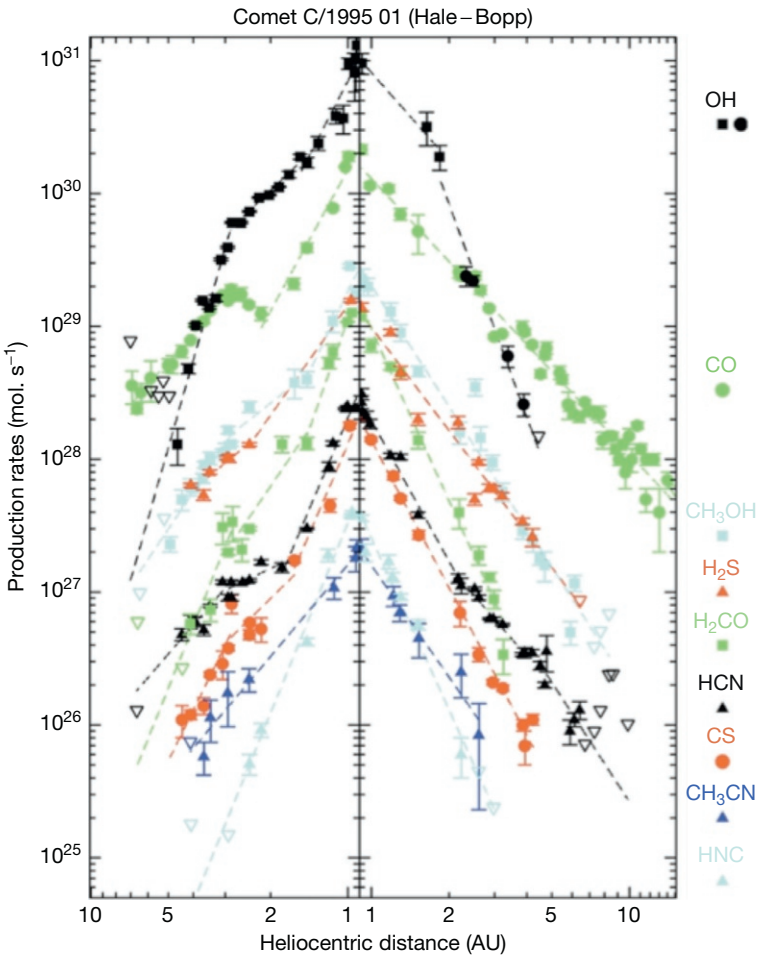


Figure 24 The so-called Christmas tree diagram, that is, the production rates from comet C/1995 O1 (Hale–Bopp) as observed using radio telescopes over the course of several years in the late 1990s and early 2000s. This was the first time that comet scientists were able to follow the production rates of so many species for such a long time interval, monitoring how their relative abundances changed as the thermal wave from the Sun penetrated to the interior and as various ices sublimated. Reproduced from the work by Biver N, Bockelée-Morvan D, Colom P, et al. (2002) The 1995–2002 long-term monitoring of comet C/1995 O1 (HALE BOPP) at radio wavelength. *Earth Moon and Planets* 90: 5–14, Figure 5, with kind permission from Springer Science and Business Media.

of 14 different species, 10 of them primary, in more than 40 comets have revealed no significant groupings related to dynamic origin (Biver et al., 2002; Crovisier et al., 2009).

The late 2000s and early 2010s have seen a burst of detections of CO_2, in over 40 comets, thanks to spacecraft observations, both *in situ* and remote (e.g., Combes et al., 1988; Crovisier et al., 1997; Feaga et al., 2007; Ootsubo et al., 2010; Reach et al., 2013). The abundance of CO_2 relative to water varies from ~3% to 30% when measured within 2.5 AU of the Sun (i.e., where water-driven outgassing is most vigorous), and no abundance variation based on primordial origin is observed (Mumma and Charnley, 2011).

Many of the organic species shown in **Figure 26** are seen, as mentioned, at infrared wavelengths through the use of high-resolution, cross-dispersed echelle spectrometers. Such instruments can observe the vibrational transitions of many species at once. This has the advantage of there being no uncertainties in relative abundances due to rotational variation. In the infrared, water, the species to which all others are compared, can also be measured directly. Species included so far in surveys are H_2O, CO, HCN, CH_3OH, H_2CO, CH_4, C_2H_2, C_2H_6, OCS, and NH_3 (Bockelée-Morvan et al, 2004; DiSanti and Mumma, 2008; Mumma et al, 2003). These surveys, which have included 26 comets so far, find three groupings based on organic abundances: organic-normal, organic-enriched, and organic-depleted. Interestingly, unlike the optical photofragment-based surveys mentioned earlier, there is no clear relation between composition and origin, since comets of both Kuiper Belt and Oort cloud origin occur in all groups. Work to increase the sample size and improve statistics is ongoing (Mumma and Charnley, 2011).

The isotopic ratios of a number of elements in comets are of interest because they are considered signatures of the environment in which the Sun formed. The D/H ratio is one of the most sought after due to the question of how much comets contributed to the Earth's water. The six Oort cloud comets measured before 2010 all had D/H ratios much higher than Standard Mean Ocean Water (SMOW). The first measurement of a Kuiper Belt comet was made by the Herschel space telescope, and P/Hartley 2 was observed to have the same

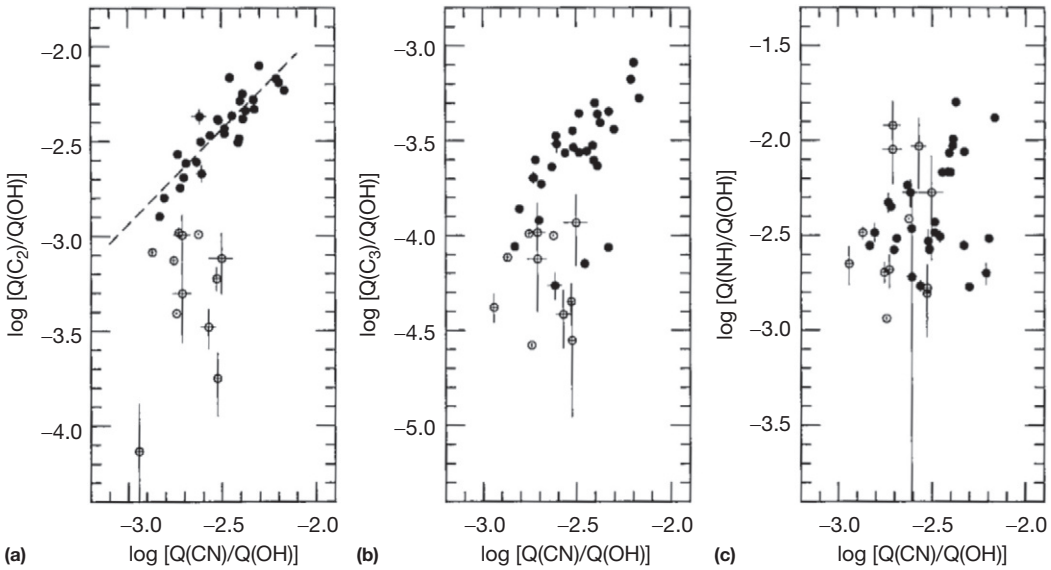


Figure 25 Correlations between relative abundances of daughter species in the comae of comets. $Q(X)$ refers to the production rate of species X. The strong depletion of C_2 and C_3 in some comets – denoted by the open circles – indicates a significant heterogeneity in the comet population. For the 'typical' comets (filled circles), the abundance of C_2 and CN (relative to water) is about the same (left panel), with the abundance of C_3 about a factor of 10 lower (middle panel). The dichotomy of abundance ratios does not extend however to NH (right panel). Reprinted from A'Hearn MF, Millis RL, Schleicher DG, Osip DJ, and Birch PV (1995) The ensemble properties of comets: Results from narrowband photometry of 85 comets, 1976–1992. *Icarus* 118: 223–270, Figure 10, with permission from Elsevier.

D/H ratio as SMOW to within the errors (Hartogh et al., 2011). This detection was followed by an upper limit for another JFC, 45P/Honda–Mrkos–Pajdušáková, which was consistent with the low D/H observed in P/Hartley 2 (Lis et al., 2013). The enigma of these observations is that the deuterium in comets originating in the Kuiper Belt was expected to be enriched relative to that of the isotropic comet population (Kavelaars et al., 2011). If the protosolar nebula began with enriched D/H similar to dense molecular clouds ($\sim 1 \times 10^{-3}$, Butner et al., 2007), and the reactions that reduce that ratio in the protoplanetary disk decrease with distance due to decreasing temperature, density, and turbulent mixing with distance (e.g., Aikawa and Herbst, 2001; Drouart et al., 1999; Mousis et al., 2000), then it would be expected that the D/H ratio would increase with increasing distance. While this led Hartogh et al. (2011) to conclude that it was possible that either these models could be incorrect or P/Hartley 2 could be an escaped Trojan (since Trojans possibly formed closer to the Sun where D/H ratios would be more depleted), the discovery of a second depleted JFC makes it even more important to measure the D/H ratio in a greater sampling of this population to determine whether this is a real taxonomic difference and also whether this can constrain our models of solar system formation. Additionally, this question has significant implications for the delivery of water to the inner solar system. A thorough review of D/H observations and challenges can be found in work by Mumma and Charnley (2011).

Since studies of comet composition make use of the coma and not the actual cometary nucleus itself, it is reasonable to wonder if the composition of the coma really can tell us the composition of the nucleus. More specifically, how do we know that the layer in the nucleus from which the coma gases originate is truly indicative of the interior composition? How much heterogeneity is there?

An important clue was provided by the Deep Impact spacecraft's visit to comet P/Tempel 1. The mission used an impactor to excavate a crater, yielding an opportunity to compare the composition of typical outgassing from surface layers with that of freshly excavated deeper layers as much as 20 m below the surface (Richardson et al., 2007). Measurements from both the ground and the spacecraft showed to within the uncertainties no change in abundance in most chemical species between the typical near surface outgassing and the ejecta plume from the excavated crater. Parent species CO_2 (Feaga et al, 2007), CO (Feldman et al., 2006), and HCN (DiSanti et al, 2007), as well as daughter fragments (photodissociation products of the larger parent molecules) NH, CN, C_3, CH, NH_2, and C_2 (Cochran et al, 2007), all retained the same relative abundance to water. So at the 0–20 m depth level, the degree of heterogeneity in comets – at least in comet P/Tempel 1 – seems to be low.

The topic of heterogeneity came to the forefront again after the spectacular data returned by the Deep Impact spacecraft as it flew by P/Hartley 2. Heterogeneity is demonstrated in **Figure 27**, which shows various images of the comet as constructed from the extensive infrared spectroscopy that was obtained. By looking at specific wavelengths in the spatially resolved spectra, spectral maps could be created. This particular figure is from the work of A'Hearn et al. (2011). The maps show that the sublimation of water was predominantly coming from the central region of the nucleus, as seen in the bottom left panel labeled 'H_2O vapor.' However, the reflected sunlight off the dust (top left) and the emission from organic species (middle right) are almost entirely spatially uncorrelated with

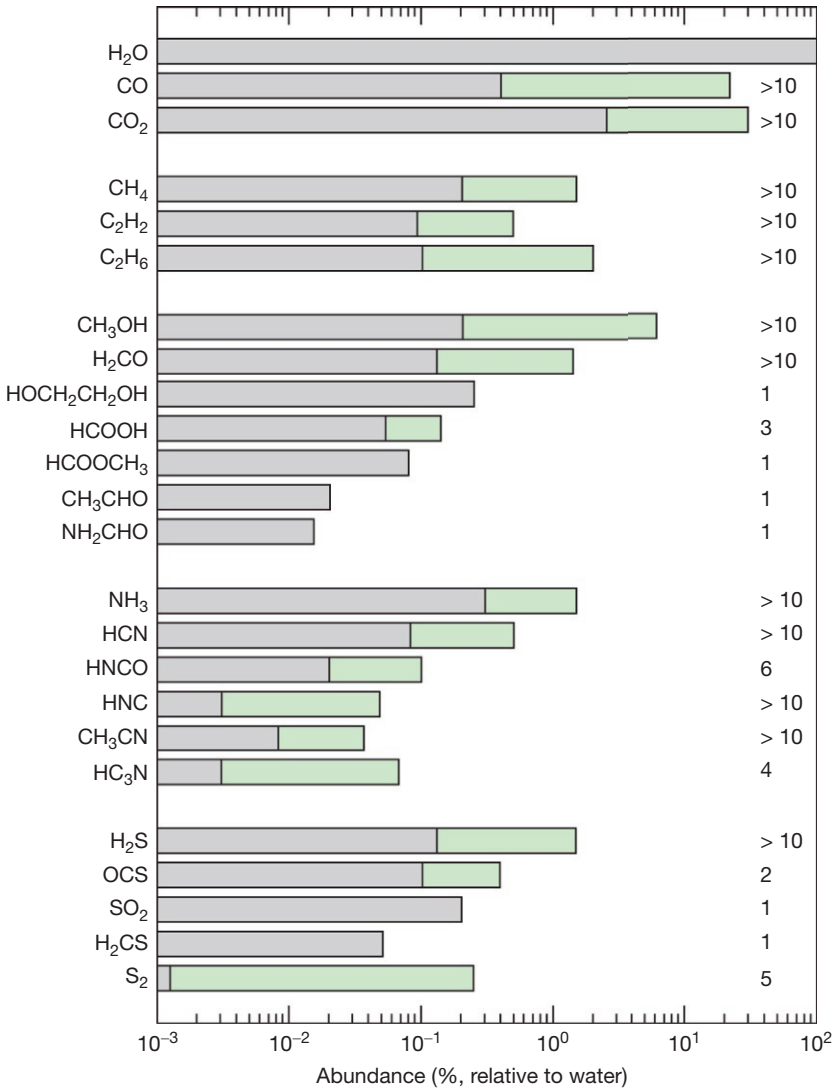


Figure 26 Relative production rates of various volatile species in comets. All abundances are scaled to water, which is set at 100%. The red segment shows the range among all measured comets; the numeral at the right edge gives how many comets in which the production rate of that species has been measured. After water, carbon dioxide, carbon monoxide, and methanol are the next most abundant species. Note the large variation in diatomic sulfur. Reprinted from Mumma MJ and Charnley SB (2011) The chemical composition of comets—emerging taxonomies and natal heritage. *Annual Review of Astronomy and Astrophysics* 49: 471–524, Figure 4, with permission from Annual Reviews.

the water vapor. Instead, they are strongly correlated with the CO_2 emission as seen in the top right panel, indicating that CO_2 at one end of the comet is driving much of P/Hartley 2's activity. Indeed, the CO_2 is even pulling off chunks of water ice (bottom right panel) into the coma.

However, the heterogeneity question is not yet fully understood, since there is good evidence that some comets are homogeneous perhaps all the way through – in particular comet P/Schwassmann–Wachmann 3, which was observed extensively in 2006 after it had broken into several large pieces during the previous two apparitions in 1995 and 2001. The gas and dust compositions in the comae of the comet's fragments were basically identical (Dello Russo et al., 2007; Sitko et al., 2011). There was no indication that the two main pieces of the comet (fragments B and C) had radically different makeup. Both comets P/Schwassmann–Wachmann 3 and P/Hartley 2 are JFCs, but the lack of heterogeneity in P/Schwassmann–Wachmann 3 suggests that the two comets could not have been formed from material that had experienced the same amount of mixing in the protoplanetary disk. Of course, there are few comets with detailed enough measurements to make a claim about heterogeneity, so this topic requires more data before broad generalizations can be justified.

The volatile content of MBCs is a new development that has taken on added importance with the claim of water ice on the surface of asteroids (24) Themis and (65) Cybele (Campins et al., 2010; Licandro et al., 2011; Rivkin and Emery, 2010). While it was certainly true that the primitive bodies in the outer main belt were thought to have appreciable water content, the observational results provide confirmation. Note however that the existence of water-bearing minerals on asteroids – as evinced by the 3 μm absorption band – has been known since the late

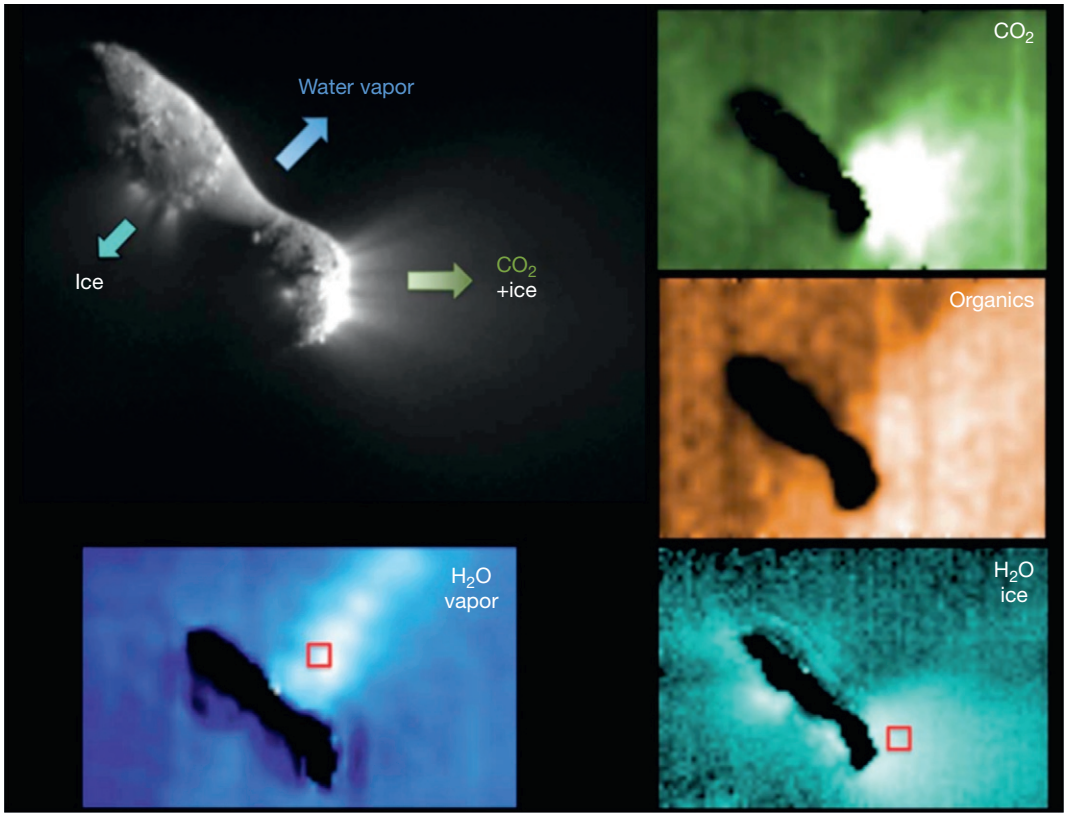


Figure 27 Heterogeneity in the nucleus of comet P/Hartley 2 as imaged by the EPOXI mission. The flyby of this comet revealed for the first time the significant role of carbon dioxide in driving cometary activity in a Jupiter-family comet in the inner solar system. Top left panel: Context image of the comet in visible-light continuum. Bottom left panel: Water vapor map, showing that most of the water is sublimating from the center of the comet's nucleus. Top right panel: CO_2 vapor map, showing that most of the carbon dioxide is sublimating from one of the ends of the nucleus. Middle right and bottom right panels: Maps of organic emission (C–H stretching bonds) and water ice absorption, which clearly show the vastly stronger correlation with the CO_2-driven activity rather than the water-driven activity. The long axis of the nucleus is about 2.3 km. Reproduced from A'Hearn MF, Belton MJS, Delamere WA, et al. (2011) EPOXI at Comet Hartley 2. *Science* 332: 1396, Figure 5, with permission from AAAS.

1970s, when it was first seen on Ceres by Lebofsky (1978). Interestingly, there is evidence of transient or variable OH coming from Ceres, first reported by A'Hearn and Feldman (1992) using IUE and then later by Küppers et al. (2014) using Herschel. In particular, the Herschel observations have shown evidence of OH while Ceres is near perihelion, and the amount is seen to vary with Ceres's rotation, suggesting specific source regions on the surface. Detection of water and/or ice in the outer part of the asteroid belt has also motivated some theoretical work on understanding how volatiles – in particular, water ice – can survive in the asteroid belt for billions of years (e.g., Schorghofer, 2008). The key seems to be having low thermal inertia so that diurnal and annual heat pulses from sunlight simply cannot make it down deep into the interior of the bodies. Thus, the natal water ice can survive for long periods of time and is available when, for example, an impactor strikes the body, clearing away the topmost layers and exposing the underlying ice to space and to the telescopes of astronomers.

The longevity of water ice and the poor thermal coupling between volatiles and rocks are also exemplified by the findings of the Deep Impact spacecraft on both comets P/Tempel 1 and P/Hartley 2, where surface ice was seen to cover a small percentage of each nucleus's surface in locations where ostensibly the local temperature of the rock was well above the water ice sublimation temperature (Groussin et al., 2013; Sunshine et al., 2006). While this ice is likely not primordial, but probably a recondensation of settled water that was formerly in the gas phase, it is remarkable that cold ice could so clearly be seen surrounded by much hotter rock. Furthermore, the fact that CO_2 was so abundant in as small a comet as P/Hartley 2 – one might have expected such a highly volatile species to have been completely baked out of the comet – lends further credence to the idea of cometary ice being decoupled from cometary silicates, probably helped by high porosity.

Much of the water content of the asteroids is studied through its existence within hydrated minerals, largely phyllosilicates. Studying the hydration is important for understanding what was going on in the young solar system such that liquid water could actually be in the physical presence of some of these minerals. In particular, the absorption band near 3 μm has been studied extensively in many asteroids, as mentioned earlier, going back over 30 years (e.g., Lebofsky, 1980). The aqueous alteration of minerals is seen in many primitive asteroids, but it is important to note that not all of them have it (e.g., Barucci et al., 1998; Rivkin, 2012; Rivkin et al., 2002). Interestingly, some of the asteroids that show evidence of hydration – for example, the so-called M types (under the Tholen taxonomy) – were once thought to be largely metallic

asteroids and to have been heated to quite high temperatures. The presence of hydrated minerals and the lack of FeO seem to contradict this scenario. Radar observations do not support most of the M types having significant iron metal, at least near the surface (Shepard et al., 2010, submitted for publication). However, iron–nickel meteorites must come from somewhere, and the parent bodies of these are still under debate (e.g., Emery and Lim, 2011). In the Bus–DeMeo taxonomy, the former M types now fall within the X-complex, distributed among several taxa, and are apparently a compositionally heterogeneous group.

10.15.5.3 Mineralogy

As stated in **Section 10.15.1.5**, **Figure 2** has a sampling of the spectral diversity among the asteroids. Note that the figure is not exhaustive; the overall slope of an asteroid's spectrum can change depending on how much weathering it has suffered, so in some cases, a single type of asteroid can actually encompass a range of spectral slopes.

In any case, the figure demonstrates that many asteroids have absorption features that are diagnostic of the surface composition; the most prominent and commonly seen ones are the deep features near 1.0 and 2.0 μm, commonly referred to as 'Band I' and 'Band II,' respectively. Both absorption bands are due to pyroxene, while olivine also contributes to Band I. The band centers and band areas can be used to place some constraints on the mineral abundances of Ca and Fe in the pyroxenes, of clino- versus orthopyroxene, and of olivine (e.g., Gaffey et al., 2002). However, it is important to note that the mixing of various different particle sizes and compositions can lead to nonunique interpretations of observed spectra.

Another frequently seen absorption feature occurs near 0.7 μm, which seems to arise in asteroids with phyllosilicates on the surface (Vilas, 1994). The oxidized iron within the hydrated mineral undergoes a charge transfer from doubly to triply ionized (i.e., Fe^{2+} to Fe^{3+}). As might be expected, the existence of this feature is correlated with an absorption feature near 3 μm that is also indicative of hydrated minerals (e.g., Rivkin, 2012; Rivkin et al., 2002); the visible-wavelength feature however is much easier to observe.

Direct sampling of asteroid mineralogy has now become possible with the success of the Hayabusa mission, which returned samples from the surface of Itokawa. In particular, one major result was that the strongly suspected link between S-type asteroids and OCs was confirmed (Nakamura et al., 2011). Furthermore, the isotopic oxygen abundance suggests that Itokawa itself is one of the sources of the OCs (Yurimoto et al., 2011).

Comet mineralogy has largely been studied through the dust grains that are dragged into space by the sublimating gas. An excellent recent review has been given by Kelley and Wooden (2009). Typical grain sizes are on the order of 1 μm, which is small enough that emission features from silicates appear in the 10 μm region, a wavelength regime amenable to ground-based spectroscopy. One of the most intriguing results from the study of many comets is the existence of crystalline silicates. Such crystallinity is in contrast to interstellar silicates, which may start out crystalline but are amorphized on million-year timescales (e.g., Kemper et al., 2004). Thus, crystalline silicates within the comets must have been heated to crystallize before being incorporated into the comet. We discuss this further in **Section 10.15.6.3**.

The silicates themselves are most commonly olivines and pyroxenes, as evinced by the shape of the 10 μm silicate feature. The shape also can be used to derive the degree of crystallinity and also the relative amounts of Mg and Fe. Many comets seem to have Mg-rich olivines and pyroxenes, or at least roughly equal amounts of Mg and Fe. This is not universal, however, and among the comets in which it has been studied, there appears to be diversity in the Mg-to-Fe silicate abundance. (Pyroxenes with some fraction of calcium (e.g., diopside and hedenbergite) are not commonly seen in comets although there are some claimed detections, including in the Deep Impact ejecta.) This high average Mg-to-Fe ratio is interesting since in asteroids and meteorites, it is more common to see the Fe-rich silicates. This dichotomy is presumably related to the compositional gradients in the protoplanetary disk. The cometary silicate composition is also somewhat dependent on dynamic class, perhaps again revealing something about formation regions of these objects.

The existence of hydrated minerals in comets, as claimed in the Deep Impact ejecta, would have significant implications for cometary formation, since presumably, the hydration could only come from liquid water, that is, water heated out of the solid phase. Yet we know that comets must have retained most of their water ice and indeed even the more volatile and highly abundant species like CO and CO_2.

Our understanding of the mineralogy of comets was helped greatly by the success of the Stardust mission, which brought dust grains from comet P/Wild 2 back to Earth. A summary of Stardust's major results in this regard has been provided by Bradley et al. (2009), and implications of the results have been reviewed by Wooden (2008). One major result from the enabled laboratory studies is that the high degree of crystallinity in the silicates – mentioned in the previous text as a result of telescopic observations – was confirmed. This was a direct laboratory measurement and so a nice corroboration of the telescopic data. Such materials must have been heated and then brought out past the snow line, potentially many AU away, in order to be incorporated into the icy comets.

The result is even more significant since it appears that the fraction within the Stardust sample of so-called 'presolar' grains – grains that condensed in the interstellar medium or around other stars before finding their way into our own protoplanetary disk – is lower than what was expected based on what is seen in primitive meteorites (Stadermann et al., 2008). At face value, this implies that the rocky component of comets is quite heavily derived from material that was in the inner solar system and flowing outward. However, there is some evidence that presolar grains were preferentially destroyed by the collection process (Floss et al., 2013), which may complicate the interpretation.

Yet another interesting aspect to the Stardust sample is the low abundance of chondrules and phyllosilicates (Nakamura et al., 2008; Zolensky et al., 2006). With the disk evidently cycling material outward to the comet-formation region fairly efficiently, one might expect that collisions among young asteroidal bodies in the inner solar system could deliver chondrules and aqueously altered materials to the region where comets

could incorporate them. Chondrules and phyllosilicates are after all common, judging by the meteorite record and by asteroid spectroscopy. However, the P/Wild 2 grains contain few of either, suggesting that the parent bodies to today's cometary nuclei are 'preaccretionary,' that is, formed before the epoch of chondrule formation (Wooden, 2008). While the details of the thermodynamics, transport, and chemistry in the protoplanetary disk are important for getting the right answer, the result can place constraints on disk evolution models.

10.15.6 Big Questions and the Future

In this last section, we summarize some of the significant problems with regard to small bodies that we think are the most pressing. We also provide some discussion of how such problems could be addressed. The study of small bodies has been recognized by the Planetary Science Decadal Survey (Committee on the Planetary Science Decadal Survey 2011) as a way to address some of the questions of the highest overarching importance in the field. What the small bodies lack in mass relative to other parts of the solar system, they make up for in importance to understanding solar system history and evolution.

There are several general points that are useful to keep in mind regarding observations of small bodies. (1) Discovery far outpaces characterization. There are over 600 000 known asteroids, almost 500 JFCs, and a few thousand known Halley-type and LPCs. It is far easier to discover such objects than to study them and discover their compositional, structural, and physical properties. (2) Many small bodies have no or only very weak spectroscopic features that can be used for compositional diagnostics. To put it another way, the jump from reflectance and thermal emission properties and taxonomy to composition and mineralogy is often not straightforward, nor unique. (3) Observations rarely sample an object deeper than a few microns; many datasets obtained on small bodies really are only addressing the topmost microscopic layer. This is especially unfortunate because it is that layer that is the most processed. Seeing deeper down into an object is very useful but often very difficult to do unambiguously. (4) While we know that the small-body populations are a diverse group, we have only begun to understand the full variety of comet and asteroid properties. In other words, we know we are ignorant, but we are not yet quite sure just how ignorant! (5) The meteorite collection is indeed vast, but a significant barrier to advancing our understanding of small bodies is the relative lack of samples with known context. At the moment, we only have samples from Itokawa and from P/Wild 2.

10.15.6.1 Where Is the Water?

As discussed earlier, we know that water makes up a significant fraction of the mass of a comet, but its contribution to the masses of the asteroids is much more uncertain. An important clue is the fact that CI and CM meteorites can be up to 15% water by weight (Salisbury et al., 1991). The water content of asteroids is a vital question since it is important to know what kind of objects actually provided Earth with its water. Part of this investigation requires studying how much of the water is bound up in hydration of minerals. This is a fundamental chemistry problem and needs to be addressed if we are to understand how much water there actually was available in the protoplanetary disk. Recent discoveries of water, or at least OH, in somewhat nonintuitive locations such as the poles of Mercury and the surface of the Moon suggest that it is crucial to know the details of the thermal environment in order to understand where water can survive.

Observationally, one approach that could vastly improve the state of the art is to have a large spectroscopic census of the 3 μm region in many asteroids. This wavelength region contains a broad absorption feature due to overlapping vibrational overtones of water and OH. The spectrum of each object should have high enough signal-to-noise to be able to use the shape of the absorption feature to distinguish between water itself and water that is bound in a hydrated mineral. Such a census would let us find surface water abundance as a function of taxonomic type and heliocentric distance and thus give us some connection between water and the birth location of the object. To some extent, this project is already underway, but spectroscopy at these wavelengths is not trivial, requiring dry air with low precipitable water above the telescope. Preliminary results suggest that the distribution of hydrated asteroids is not related to asteroid size or orbital location in a simple way (e.g., Rivkin et al., 2002).

10.15.6.2 Where Are the Organics?

As with water, the key to understanding the organics is in spectroscopic studies of the 3.2–3.6 μm region. C–H bonds in organic species have stretching modes at those wavelengths. High signal-to-noise is needed to be able to differentiate between different sets of organics within an asteroid. Currently, one usually just sees a broad feature that can be difficult to assign to specific species, since the phase of the material is solid. Organic materials in meteorites have been studied extensively, but the relationship to specific asteroids or locations is lacking (e.g., Pizzarello et al., 2006 and references therein).

Fortunately, comets can provide more insight about organic abundances, for the icy bodies at least, since the cometary activity naturally dredges up interior volatiles, making them visible in the coma as discussed earlier in this chapter. The difficulty is determining which volatiles are native to the cometary nucleus itself and which are products of chemistry that can occur within the collision zone of the coma, within a few radii of the nucleus's surface.

There has been a tremendous increase in the study of cometary organics since the late 1990s, owing to improvements in infrared spectrometers and to access to the large telescopes that are required. Thanks to these observations, a clearer picture is emerging of how various species – C_2H_2, CH_3OH, C_2H_6, H_2CO, etc. – are distributed among the comet population (Mumma and Charnley, 2011). Continued observation of comets – from all dynamic classes – is certainly warranted. The number statistics are still fairly low, so a census of the organics in even 10–20 more comets, both short and long period, would be very helpful to discern trends. The difficulty is that there are only a few comets per year that are bright

enough for this kind of observation. This suggests that JWST, representing potentially a tremendous jump forward in the number of faint targets that are accessible, could bring considerable data to the problem.

A sample return mission to a comet could in principle provide us with cometary species that could be studied in exquisite detail in the laboratory. This would be a boon for our understanding of planetesimal accretion and cometary evolution. Such a mission would be technically complex, however, since one problem to be solved is preserving organics and volatiles at low temperatures during the journey back to Earth so that no chemistry occurs along the way. The Stardust mission unfortunately was not able to bring back many organics since most of the organics sublimated away as the dust grains hit the collector aerogel at high relative speed. One interesting result however was the discovery of glycine in some of the Stardust samples (Elsila et al., 2009). We can look forward to the day when more amino acids and other complex species will be measured in cometary materials.

10.15.6.3 What Was the Thermal and Compositional Gradient in the Protoplanetary Disk? How Much Mixing Was There?

An apparent compositional gradient in the current solar system has been identified for several decades (Carvano et al., 2010; Gradie et al., 1989), as was described in **Section 10.15.2**. Recent dynamic modeling – in particular models like the Nice model and the Grand Tack model that have the giant planets migrating significantly and disrupting the orbital distribution of planetesimals – has suggested that the interpretation of this apparent, current gradient is not easy. The goal, after all, is to find the *primordial* compositional gradient. Two main problems in understanding what the primordial gradients were like remain: (a) there is uncertainty in what dynamic mixing has happened, and (b) we do not have enough compositional data to begin with. Detailed compositional studies of a wide range of small bodies will be necessary if we are to address the situation, and naturally, the first step is understanding composition as a function of object type and object's location in the solar system right now. We can then hope to make use of advances in our understanding of solar system dynamics to constrain the initial conditions.

Although it had been suspected earlier, strong evidence for tremendous mixing in the protoplanetary disk comes from the fact that Stardust brought back silicate dust grains from comet P/Wild 2 with a degree of crystallinity that could only have been formed in high temperatures (Zolensky et al., 2006). This result was discussed in **Section 10.15.5.3**. In fact, more than half of the Fe-bearing silicates show evidence of having been processed in the inner solar system (Ogliore et al., 2009).

There are several models that could explain the transport of small heated grains to the outer solar system. The Shu X-wind model (Shu et al., 1994, 2000) has been a leading hypothesis since it can explain astrophysical phenomena regarding disks and jets around young stellar objects. The model uses the magnetohydrodynamics of the star–disk interaction to fling material outward after it has had a chance to be heated close to the protostar. The magnetosphere of the accreting star transfers angular momentum to the disk, which induces an outward wind. While much of the wind blows along field lines out of the disk altogether, some grains that were heated in the stellar accretion region will be blown back into the disk at distances appropriate for cometary accretion. However, there are several significant problems with this scenario as the source of crystalline grains, including the possible lack of time to produce the inferred initial $^{26}Al/^{27}Al$ ratio (Boss, 2012; Desch et al., 2010).

Other possible explanations come from one- and two-dimensional hydrodynamic models of turbulent, viscous disks. These models explain the outward transport of grains through diffusion and/or midplane gas flows in the first 1–2 My of disk formation. These grains would have to be incorporated into icy bodies very early on, before they fell back to smaller radii (e.g., Hughes and Armitage, 2010). Additionally, in full 3-D models of 'marginally gravitationally unstable' (MGU) disks, the disk appears to evolve through one or more phases of gravitational instability, and self-gravitational torques can transport material both inward and outward (e.g., Boss, 2007). Each of these approaches has implications for mixing and heterogeneity in the disk at varying timescales.

In any case, one of the significant clues to the early protoplanetary disk lies in the solids with depleted initial ^{26}Al abundance. In meteorites, both chondrules and a fraction of the calcium–aluminum-rich inclusions show evidence for zero initial abundance, indicating they formed before or several million years after the event(s) that infused the disk with ^{26}Al. According to Boss (2012), better Pb–Pb isochron ages for the depleted inclusions, as well as tighter constraints on the initial heterogeneity of short-lived isotopes ^{60}Fe and ^{26}Al, will help us understand the initial conditions. Furthermore, measurements of the isotopic ratio of oxygen in molecules observed in protoplanetary disks will help us understand mechanisms for fractionation and spatial heterogeneity of oxygen isotopic abundances since these isotopes display evidence for mass-independent fractionation in our solar system (Boss, 2012; Clayton, 1993; Sakamoto et al., 2007).

While meteorites continue to be a source of samples, sample return missions represent a promising way to make further progress, since parent bodies can be chosen for their known composition and dynamic properties.

10.15.6.4 What Did Small Bodies Do to Earth?

The current dominant hypothesis is that Earth and the terrestrial planets likely formed too close to the young Sun in too warm a region to allow hydration at formation (Morbidelli et al., 2000). In simplistic terms, this was thought to indicate that there was a 'frost line' or 'snow line' in the protoplanetary disk that marked the boundary of where water could condense out of the gas phase. Recent work has shown this is indeed an oversimplification and that condensation in the real protoplanetary disk was more complicated, as well as time-dependent (e.g., Bell, 2010; Encrenaz, 2008; Morbidelli et al., 2012; van Dishoeck et al., 2014). While it has long been assumed that small bodies from the outer solar system brought water and organics to what would have otherwise been a dry Earth (e.g., Matsui and Abe, 1986), the lunar D/H ratio and the Earth's mantle composition both suggest that it was *possible* for there to have been enough water inside the snow line for Earth to have formed wet and thus that very little water was brought

from impactors (e.g., Marty, 2012; Saal et al., 2013). On the other hand, the composition and isotopic ratios in the mantle have also been used to argue in favor of contributions by both asteroids and comets (e.g., Halliday, 2013). Greater knowledge of isotopic ratios of small bodies and of the history of fractionation in the solar system will be a key part of understanding the origin of the Earth's water.

As discussed in Section 10.15.5.2, until 2010, the water D/H ratio in six Oort cloud comets showed that the D/H ratio was much higher than observed in the Earth's oceans (i.e., in SMOW). This made having comets as a source of the Earth's water problematic. Modeling suggested that if Jupiter were in a nearly circular orbit, then water could be delivered to the terrestrial planets by chondritic material from beyond the frost line (Raymond et al., 2004). The discovery that the JFC P/Hartley 2 has a water D/H ratio similar to SMOW (Hartogh et al., 2011) revives the possibility that comets could have delivered a significant fraction of the Earth's water. Furthermore, two other recent results on the D/H ratio – (a) the D/H measurement of less than 1.5 times SMOW in the Oort cloud comet C/2009 P1 (Garradd) by Bockelée-Morvan et al. (2012) and (b) the finding of an upper limit to the D/H ratio in comet P/Honda–Mrkos–Pajdušáková that is consistent with that of P/Hartley 2 (Lis et al., 2013) – demonstrate a possible continuum in cometary D/H. Clearly, the D/H ratio in comets is a more complicated issue than was previously thought. In any case, given the success of the asteroid models at delivering water, the answer may be that the Earth's volatiles were delivered by some combination of asteroids and comets – unless it turns out that the wet Earth hypothesis is correct. It is possible that we will need to develop and fly the technology to sample small-body noble gas isotopes before we are reasonably certain of the origin of the Earth's water.

It is clear from both this question and several of the others posed in this section that understanding the dynamic evolution of the small bodies is critical to being able to reconstruct the protoplanetary disk. The current dominant model of large-scale planetary migration in our solar system, the Nice model mentioned in earlier sections, has done an admirable job of explaining many observed phenomena, but the difficult part is using it to make a prediction that both (a) is observationally feasible and (b) would make a strong test of the idea's validity. Compositional information on the small bodies is one possible way – since, for example, the model would predict that one would see a similar primordial composition among the outer main-belt asteroids, Trojans, and trans-Neptunian objects. The difficulty is that it is still hard to interpret the surface composition due to a lack of diagnostic features. What is unknown is whether the surfaces would really be the same because of similar primordial condition or simply because of similar subsequent evolution. Sample return missions may be required before we know that answer.

Nongravitational forces are an important aspect to the dynamics of many objects, as it is now clear that over fairly short dynamic timescales, they can move asteroids and comets significantly around the solar system. In particular, for asteroids, the Yarkovsky and YORP effects need to be better understood – and this means understanding the thermal properties as well as the structural properties in the interior. This is important since the Yarkovsky effect affects the delivery of asteroids from the asteroid belt to near-Earth space, including those that could someday strike Earth. For comets, outgassing torques and forces – on the whole much stronger than Yarkovsky effects – can measurably change a comet's orbital elements on month- or year-long timescales, thereby making it more difficult to gauge just how significant the cometary impact risk is for Earth.

10.15.6.5 Forthcoming Breakthroughs

The next decade should bring tremendous insight into our understanding of comets and asteroids. First, the Rosetta spacecraft will rendezvous with comet P/Churyumov–Gerasimenko in 2014 and observe it close-up for several months. Rosetta is primarily a European Space Agency mission but NASA has contributed some of its instruments. For the first time, we will be able to watch a comet turn on as its activity gets started. The spacecraft also has a lander, Philae, that will be able to study the surface directly, our first such opportunity to do so on a comet. The spacecraft will arrive at comet P/Churyumov–Gerasimenko in May 2014 – when the comet is 4.0 AU from the Sun, outside the zone where water could sublimate – and will stay with the comet at least until December 2015. The comet's perihelion occurs in August 2015, when it will be 1.24 AU from the Sun.

Asteroid sample return should figure prominently in the next several years, with the United States, Europe, and Japan all potentially sending sample return missions to primitive asteroids. The Japanese mission, Hayabusa 2, is a follow-up to its successful Hayabusa mission that brought back samples from asteroid Itokawa in 2010. Hayabusa 2 will travel to (162173) 1999 JU3, which is a primitive asteroid. Launch would happen in 2014 or 2015, with asteroid arrival in 2018. The spacecraft would collect a sample and leave the asteroid in 2019, returning to Earth in 2020. The US mission, OSIRIS-REx, which stands for 'Origins Spectral Interpretation Resource Identification Security Regolith Explorer,' will retrieve samples from asteroid Bennu. This object is likewise a primitive asteroid. Launch will happen in 2016, with arrival to the asteroid in 2019. The spacecraft would study the asteroid for many months, departing likely in 2021 and getting back to Earth in 2023. Even if only one of these missions is successful, we can expect to have tremendous new insights into our understanding of primitive bodies in the 2020s.

10.15.6.6 Longer-Term Missions

In the farther future, significant scientific return can be obtained from robotic missions that visit Trojan asteroids, primitive main-belt asteroids, and dormant/extinct comets. In the United States, NASA has plans for human visitation of asteroids in the relatively near future, mid-2020s. However, it is unclear exactly what path toward that goal will be taken. Nonetheless, there are significant engineering obstacles to a successful manned mission to an asteroid that need to be overcome. Some of these obstacles – sometimes known as 'strategic knowledge gaps' – can be answered through telescopic observation and robotic visitation of the asteroids in question. Some of these questions include the following: Could astronauts actually operate in the nearly zero gravity

environment around an asteroid? How would the spacecraft rendezvous? What are the surface and immediate subsurface layers like? Can one grab onto the surface or is it too powdery? Could one anchor to it? What thermal variation would there be as the asteroid rotates in and out of sunlight? Is the asteroid in a principal-axis rotation state, or is it tumbling? How far down would one have to dig to get to more pristine material, if at all?

While these are difficult problems to solve, one cannot help but be excited about what secrets of the small bodies and of our solar system will be revealed in the coming years. The synergy of Earth-based telescopic observation (to provide the big picture about many objects) and spacecraft missions (to provide the detailed picture of a few key objects) is sure to continue to provide astronomers and geophysicists with new, thrilling insights.

Acknowledgments

We greatly appreciate the assistance of Michael Nolan and Patrick Taylor in preparing this manuscript. We also thank Peter Thomas and an anonymous reviewer for their careful reading and suggestions. Some of our figures made use of data archived by the Small Bodies Node of the NASA Planetary Data System. YRF and ESH acknowledge support from NSF (through Grant No. AST-1109855 to the Johns Hopkins University). LMW acknowledges support from NASA (through Grant No. NNX09AB71G issued through the Planetary Atmospheres program). The Arecibo Planetary Radar Program is supported by the National Aeronautics and Space Administration under Grant No. NNX12AF24G issued through the Near-Earth Object Observations Program. The Arecibo Observatory is currently operated by SRI International under a cooperative agreement with the National Science Foundation (AST-1100968) and in alliance with Ana G. Mendez-Universidad Metropolitana and the Universities Space Research Association. At the time of some of these observations, the Arecibo Observatory was operated by Cornell University under a cooperative agreement with the National Science Foundation.

References

A'Hearn MF, Belton MJS, Delamere WA, et al. (2011) EPOXI at Comet Hartley 2. *Science* 332: 1396.

A'Hearn MF and Feldman PD (1992) Water vaporization on Ceres. *Icarus* 98: 54–60.

A'Hearn MF, Millis RL, Schleicher DG, Osip DJ, and Birch PV (1995) The ensemble properties of comets: Results from narrowband photometry of 85 comets, 1976–1992. *Icarus* 118: 223–270.

Abe S, Mukai T, Hirata N, et al. (2006) Mass and local topography measurements of Itokawa by Hayabusa. *Science* 312: 1344–1349.

Aikawa Y and Herbst E (2001) Two-dimensional distributions and column densities of gaseous molecules in protoplanetary disks. II. Deuterated species and UV shielding by ambient clouds. *Astronomy and Astrophysics* 371: 1107.

Alvarez LW, Alvarez W, Asaro F, and Michel HV (1980) Extraterrestrial cause for the Cretaceous–Tertiary extinction. *Science* 208: 1095–1108.

Ammannito E, De Sanctis M, Capaccioni F, et al. (2013) Vestan lithologies mapped by the visual and infrared spectrometer on Dawn. *Meteoritics and Planetary Science* 48: 2185–2198.

Anders E (1989) Pre-biotic organic matter from comets and asteroids. *Nature* 342: 255–257.

Asphaug E (1999) Survival of the weakest. *Nature* 402: 127–128.

Barnouin-Jha OS, Cheng AF, Mukai T, et al. (2008) Small-scale topography of 25143 Itokawa from the Hayabusa laser altimeter. *Icarus* 198: 108–124.

Barucci MA, Cruikshank DP, Mottola S, and Lazzarin M (2002) Physical properties of Trojan and Centaur asteroids. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 273–287. Tucson: University of Arizona Press.

Barucci MA, Doressoundiram A, Fulchignoni M, et al. (1998) Search for aqueously altered materials on asteroids. *Icarus* 132: 388–396.

Bell JF (2010) Water on planets. *Highlights of Astronomy* 15: 29–44.

Bell JF, Fanale F, and Cruikshank DP (1993) Chemical and physical properties of the Martian satellites. In: Lewis JS, Matthews MS, and Guerrier ML (eds.) *Resources of Near-Earth Space*, pp. 887–901. Tucson: University of Arizona Press.

Belton MJS (2010) Cometary activity, active areas, and a mechanism for collimated outflows on 1P, 9P, 19P, and 81P. *Icarus* 210: 881–897.

Belton MJS, Chapman CR, Veverka J, et al. (1994) First images of asteroid 243 Ida. *Science* 265: 1543–1547.

Belton MJS, Feldman PD, A'Hearn MF, and Carcich B (2008) Cometary cryovolcanism: Source regions and a model for the UT 2005 June 14 and other mini-outbursts on Comet 9P/Tempel 1. *Icarus* 198: 189–207.

Belton MJS and Melosh J (2009) Fluidization and multiphase transport of particulate cometary material as an explanation of the smooth terrains and repetitive outbursts on 9P/Tempel 1. *Icarus* 200: 280–291.

Belton MJS, Mueller BEA, Julian WH, and Anderson AJ (1991) The spin state and homogeneity of Comet Halley's nucleus. *Icarus* 93: 183–193.

Belton MJS, Thomas P, Carcich B, et al. (2013) The origin of pits on 9P/Tempel 1 and the geologic signature of outbursts in Stardust-NExT images. *Icarus* 222: 477–486.

Belton MJS, Thomas P, Veverka J, et al. (2007) The internal structure of Jupiter family cometary nuclei from Deep Impact observations: The "talps" or "layered pile" model. *Icarus* 187: 332–344.

Belton MJS, Veverka J, Thomas P, et al. (1992) Galileo encounter with 951 Gaspra: First pictures of an asteroid. *Science* 257: 1647–1652.

Benner LAM, Nolan MC, Margot J, et al. (2008) Arecibo and Goldstone radar imaging of contact binary near-Earth asteroids. *AAS/Division for Planetary Sciences Meeting Abstracts* 40: 432.

Bertini I (2011) Main Belt Comets: A new class of small bodies in the Solar System. *Planetary and Space Science* 59: 365–377.

Besse S, Lamy P, Jorda L, Marchi S, and Barbieri C (2012) Identification and physical properties of craters on Asteroid (2867) Steins. *Icarus* 221: 1119–1129.

Bessel FW (1836) Beobachtungen über die physische Beschaffenheit des Halley'schen Kometen und dadurch veranlasste Bemerkungen. *Astronomische Nachrichten* 13: 185–232.

Binzel RP, Bus SJ, Burbine TH, and Sunshine JM (1996) Spectral properties of near-Earth asteroids: Evidence for sources of ordinary chondrite meteorites. *Science* 273: 946–948.

Binzel RP, Lupishko D, di Martino M, Whiteley RJ, and Hahn GJ (2002) Physical properties of near-Earth objects. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 255–271. Tucson: University of Arizona Press.

Binzel RP, Morbidelli A, Merouane S, et al. (2010) Earth encounters as the origin of fresh surfaces on near-Earth asteroids. *Nature* 463: 331–334.

Binzel RP, Rivkin AS, Stuart JS, et al. (2004) Observed spectral properties of near-Earth objects: results for population distribution, source regions, and space weathering processes. *Icarus* 170: 259–294.

Binzel RP and Xu S (1993) Chips off of asteroid 4 Vesta: Evidence for the parent body of basaltic achondrite meteorites. *Science* 260: 186–191.

Biver N, Bockelée-Morvan D, Colom P, et al. (1997) Long-term evolution of the outgassing of comet Hale-Bopp from radio observations. *Earth, Moon, and Planets* 78: 5–11.

Biver N, Bockelée-Morvan D, Colom P, et al. (2002) The 1995–2002 long-term monitoring of comet C/1995 O1 (HALE BOPP) at radio wavelength. *Earth, Moon, and Planets* 90: 5–14.

Bockelée-Morvan D, Biver N, Swinyard B, et al. (2012) Herschel measurements of the D/H and 16O/18O ratios in water in the Oort-cloud comet C/2009 P1 (Garradd). *Astronomy and Astrophysics* 544: L15.

Bockelée-Morvan D, Crovisier J, Mumma MJ, and Weaver HA (2004) The composition of cometary volatiles. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 391–423. Tucson: University of Arizona Press.

Borucki WJ, Koch DG, Basri G, et al. (2011) Characteristics of planetary candidates observed by Kepler. II. Analysis of the first four months of data. *Astrophysical Journal* 736: 19.

Boss AP (2007) Evolution of the solar nebula. VIII. Spatial and temporal heterogeneity of short-lived radioisotopes and stable oxygen isotopes. *Astrophysical Journal* 660: 1707–1714.

Boss AP (2012) Mixing and transport of isotopic heterogeneity in the early Solar System. *Annual Review of Earth and Planetary Sciences* 40: 23–43.

Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (2002a) An overview of the asteroids: The Asteroids III perspective. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 3–15. Tucson: University of Arizona Press.

Bottke WF Jr., Durda DD, Nesvorný D, et al. (2005) The fossilized size distribution of the main asteroid belt. *Icarus* 175: 111–140.

Bottke WF Jr., Morbidelli A, Jedicke R, et al. (2002b) Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* 156: 399–433.

Bottke WF Jr., Vokrouhlicky D, Rubincam DP, and Nesvorny D (2006) The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annual Review of Earth and Planetary Sciences* 34: 157–191.

Bradley JP, Jones AP, and Brownlee DE (2009) Insight from the unexpected. *Meteoritics and Planetary Science* 44: 1403–1405.

Bredichin T (1877) Sur les formes anomales dans le developpement des cometes. *Annales de l'Observatiore de Moscou* 1: 1–46.

Britt DT, Boice DC, Buratti BJ, et al. (2004) The morphology and surface processes of Comet 19/P Borrelly. *Icarus* 167: 45–53.

Britt DT and Consolmagno GJ (2003) Stony meteorite porosities and densities: A review of the data through 2001. *Meteoritics and Planetary Science* 38: 1161–1180.

Brown P (2013) A preliminary report on the Chelyabinsk fireball/airburst. *WGN, Journal of the International Meteor Organization* 41: 22.

Brown P, Marchenko V, Moser DE, et al. (2013) Meteorites from meteor showers: A case study of the Taurids. *Meteoritics and Planetary Science* 48: 270–288.

Brownlee DE, Horz F, Newburn RL, et al. (2004) Surface of young Jupiter family comet 81 P/Wild 2: View from the Stardust spacecraft. *Science* 304: 1764–1769.

Brunetto R, Roush TL, Marra AC, and Orofino V (2007) Optical characterization of laser ablated silicates. *Icarus* 191: 381–393.

Brunetto R and Strazzulla G (2005) Elastic collisions in ion irradiation experiments: A mechanism for space weathering of silicates. *Icarus* 179: 265–273.

Brunetto R, Vernazza P, Marchi S, et al. (2006) Modeling asteroid surfaces from observations and irradiation experiments: The case of 832 Karin. *Icarus* 184: 327–337.

Buczkowski DL, Barnouin-Jha OS, and Prockter LM (2008) 433 Eros lineaments: Global mapping and analysis. *Icarus* 193: 39–52.

Buczkowski DL, Wyrick DY, Iyer KA, et al. (2012) Large-scale troughs on Vesta: A signature of planetary tectonics. *Geophysical Research Letters* 39: L18205.

Bus SJ and Binzel RP (2002) Phase II of the Small Main-Belt Asteroid Spectroscopic Survey: A feature-based taxonomy. *Icarus* 158: 146–177.

Butner HM, Charnley SB, Ceccarelli C, et al. (2007) Discovery of interstellar heavy water. *Astrophysical Journal* 659: L137.

Campins H, Hargrove K, Pinilla-Alonso N, et al. (2010) Water ice and organics on the surface of the asteroid 24 Themis. *Nature* 464: 1320–1321.

Campins H and Swindle TD (1998) Expected characteristics of cometary meteorites. *Meteoritics and Planetary Science* 33: 1201–1211.

Carr MH, Kirk RL, McEwen A, et al. (1994) The geology of Gaspra. *Icarus* 107: 61.

Carry B (2012) Density of asteroids. *Planetary and Space Science* 73: 98–118.

Carvano JM, Hasselmann PH, Lazzaro D, and Mothé-Diniz T (2010) SDSS-based taxonomic classification and orbital distribution of main belt asteroids. *Astronomy and Astrophysics* 510: A43.

Chapman CR (1996) S-Type asteroids, ordinary chondrites, and space weathering: The evidence from Galileo's fly-bys of Gaspra and Ida. *Meteoritics and Planetary Science* 31: 699–725.

Chapman CR (1997) Gaspra and Ida: Implications of spacecraft reconnaissance for NEO issues. *Annals of the New York Academy of Sciences* 822: 227.

Chapman CR (2002) Cratering on asteroids from Galileo and NEAR Shoemaker. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 315–329. Tucson: University of Arizona Press.

Chapman CR (2004) Space weathering of asteroid surfaces. *Annual Review of Earth and Planetary Sciences* 32: 539–567.

Chapman CR, Merline WJ, Thomas PC, et al. (2002) Impact history of Eros: Craters and boulders. *Icarus* 155: 104–118.

Chapman CR, Ryan EV, Merline WJ, et al. (1996a) Cratering on Ida. *Icarus* 120: 77–86.

Chapman CR, Veverka J, Belton MJS, Neukum G, and Morrison D (1996b) Cratering on Gaspra. *Icarus* 120: 231–245.

Cheng AF and Barnouin-Jha OS (1999) Giant craters on Mathilde. *Icarus* 140: 34–48.

Cheng AF, Barnouin-Jha O, Zuber MT, et al. (2001) Laser altimetry of small-scale features on 433 Eros from NEAR-Shoemaker. *Science* 292: 488–491.

Cheng AF, Izenberg N, Chapman CR, and Zuber MT (2002) Ponded deposits on asteroid 433 Eros. *Meteoritics and Planetary Science* 37: 1095–1105.

Chesley SR, Belton MJS, Carcich B, et al. (2013) An updated rotation model for Comet 9P/Tempel 1. *Icarus* 222: 516–525.

Chiang E and Youdin AN (2010) Forming planetesimals in solar and extrasolar nebulae. *Annual Review of Earth and Planetary Sciences* 38: 493–522.

Cintala MJ, Garvin JB, and Wetzel SJ (1982) The distribution of blocks around a fresh lunar mare crater. *Lunar and Planetary Institute Science Conference Abstracts* 13: 100–101.

Clark BE, Lucey P, Helfenstein P, et al. (2001) Space weathering on Eros: Constraints from albedo and spectral measurements of Psyche crater. *Meteoritics and Planetary Science* 36: 1617–1637.

Clayton RN (1993) Oxygen isotopes in meteorites. *Annual Review of Earth and Planetary Sciences* 21: 115–149.

Cochran AL, Jackson WM, Meech KJ, and Glaz M (2007) Observations of Comet 9P/Tempel 1 with the Keck 1 HIRES instrument during Deep Impact. *Icarus* 187: 156–166.

Combes M, Crovisier J, Encrenaz T, Moroz VI, and Bibring J-P (1988) The 2.5–12 micron spectrum of Comet Halley from the IKS-VEGA experiment. *Icarus* 76: 404–436.

Committee on the Planetary Science Decadal Survey, Space Studies Board, Division on Engineering and Physical Sciences, National Research Council (2011) *Vision and Voyages for Planetary Science in the Decade 2013–2022*. Washington, DC: National Academies Press.

Coradini A, Capaccioni F, Erard S, et al. (2011) The surface composition and temperature of asteroid 21 Lutetia as observed by Rosetta/VIRTIS. *Science* 334: 492.

Crovisier J, Biver N, Bockelée-Morvan D, and Colom P (2009) Radio observations of Jupiter-family comets. *Planetary and Space Science* 57: 1162–1174.

Crovisier J, Leech K, Bockelée-Morvan D, et al. (1997) The spectrum of Comet Hale-Bopp (C/1995 01) observed with the Infrared Space Observatory at 2.9 AU from the Sun. *Science* 275: 1904–1907.

Davidsson BJR (2001) Tidal splitting and rotational breakup of solid biaxial ellipsoids. *Icarus* 149: 375–383.

Davidsson BJR and Gutiérrez PJ (2004) Estimating the nucleus density of Comet 19P/Borrelly. *Icarus* 168: 392–408.

Davidsson BJR and Gutiérrez PJ (2006) Non-gravitational force modeling of Comet 81P/Wild 2. I. A nucleus bulk density estimate. *Icarus* 180: 224–242.

Davis DR (1999) The collisional history of asteroid 253 Mathilde. *Icarus* 140: 49–52.

Davis DR, Chapman CR, Durda DD, Farinella P, and Marzari F (1996) The formation and collisional/dynamical evolution of the Ida/Dactyl system as part of the Koronis family. *Icarus* 120: 220–230.

Davis DR, Durda DD, Marzari F, Campo Bagatin A, and Gil-Hutton R (2002) Collisional evolution of small-body populations. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 545–558. Tucson: University of Arizona Press.

De Sanctis MC, Ammannito E, Capria MT, et al. (2012) Spectroscopic characterization of mineralogy and its diversity across Vesta. *Science* 336: 697.

De Sanctis MC, Ammannito E, Capria MT, et al. (2013) Vesta's mineralogical composition as revealed by the visible and infrared spectrometer on Dawn. *Meteoritics and Planetary Science* 48: 2166–2184.

Dello Russo N, Vervack RJ, Weaver HA, et al. (2007) Compositional homogeneity in the fragmented comet 73P/Schwassmann-Wachmann 3. *Nature* 448: 172–175.

DeMeo F and Binzel RP (2008) Comets in the near-Earth object population. *Icarus* 194: 436–449.

DeMeo FE, Binzel RP, Slivan SM, and Bus SJ (2009) An extension of the Bus asteroid taxonomy into the near-infrared. *Icarus* 202: 160–180.

Demura H, Kobayashi S, Nemoto E, et al. (2006) Pole and global shape of 25143 Itokawa. *Science* 312: 1347–1349.

Denevi BW, Coman EI, Carsenty U, et al. (2012) Global variations in regolith depth on asteroid Vesta. European Planetary Science Congress 2012 EPSC2012.

Descamps P, Marchis F, Berthier J, et al. (2011) Triplicity and physical characteristics of Asteroid (216) Kleopatra. *Icarus* 211: 1022–1033.

Desch SJ, Morris MA, Connolly HC Jr., and Boss AP (2010) A critical examination of the X-wind model for chondrule and calcium-rich, aluminium-rich inclusion formation and radionuclide production. *Astrophysical Journal* 725: 692–711.

DiSanti MA and Mumma MJ (2008) Reservoirs for comets: Compositional differences based on infrared observations. *Space Science Reviews* 138: 127–145.

DiSanti MA, Villanueva GL, Bonev BP, et al. (2007) Temporal evolution of parent volatiles and dust in Comet 9P/Tempel 1 resulting from the Deep Impact experiment. *Icarus* 191: 481–493.

Drake MJ (2001) The eucrite/Vesta story. *Meteoritics and Planetary Science* 36: 501–513.

Drake MJ and Campins H (2006) Origin of water on the terrestrial planets. In: Daniela L, Sylvio Ferraz M, and Angel FJ (eds.) *Asteroids, Comets, Meteors, Vol. 229: IAU Symposium*, pp. 381–394. Cambridge: Cambridge University Press.

Drouart A, Dubrulle B, Gautier D, and Robert F (1999) Structure and transport in the solar nebula from constraints on deuterium enrichment and giant planets formation. *Icarus* 140: 129.

Dukes CA, Baragiola RA, and McFadden LA (1999) Surface modification of olivine by H^+ and He^+ bombardment. *Journal of Geophysical Research* 104: 1865–1872.

Duncan MJ and Levison HF (1997) A scattered comet disk and the origin of Jupiter family comets. *Science* 276: 1670–1672.

Duncan M, Levison H, and Dones L (2004) Dynamical evolution of ecliptic comets. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 193–204. Tucson: University of Arizona Press.

Elsila JE, Glavin DP, and Dworkin JP (2009) Cometary glycine detected in samples returned by Stardust. *Meteoritics and Planetary Science* 44: 1323–1330.
Emery JP and Lim LF (2011) Thermal emission spectroscopy of M-type asteroids: How many are metallic? *EPSC-DPS Joint Meeting 2011*, p. 1438.
Encrenaz T (2008) Water in the Solar System. *Annual Review of Astronomy and Astrophysics* 46: 57–87.
Farinella P and Davis DR (1996) Short-period comets: Primordial bodies or collisional fragments? *Science* 273: 938–941.
Farnham TL, Bodewits D, Li J-Y, et al. (2013) Connections between the jet activity and surface features on Comet 9P/Tempel 1. *Icarus* 222: 540–549.
Feaga LM, A'Hearn MF, Sunshine JM, Groussin O, and Farnham TL (2007) Asymmetries in the distribution of H_2O and CO_2 in the inner coma of Comet 9P/Tempel 1 as observed by Deep Impact. *Icarus* 190: 345–356.
Feldman PD, Lupu RE, McCandliss SR, et al. (2006) Carbon monoxide in Comet 9P/Tempel 1 before and after the Deep Impact encounter. *The Astrophysical Journal Letters* 647: L61–L64.
Fernández YR, Jewitt DC, and Sheppard SS (2005) Albedos of asteroids in comet-like orbits. *Astronomical Journal* 130: 308–318.
Fernández YR, Kelley MS, Lamy PL, et al. (2013) Thermal properties, sizes, and size distribution of Jupiter-family cometary nuclei. *Icarus* 226: 1138–1170.
Festou MC, Keller HU, and Weaver HA (2004) A brief conceptual history of cometary science. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 3–16. Tucson: University of Arizona Press.
Fink U (2009) A taxonomic survey of comet composition 1985–2004 using CCD spectroscopy. *Icarus* 201: 311–334.
Floss C, Stadermann FJ, Kearsley AT, et al. (2013) The abundance of presolar grains in comet 81P/Wild 2. *Astrophysical Journal* 763: 140.
Foderà Serio G, Manara A, and Sicoli P (2002) Giuseppe Piazzi and the discovery of Ceres. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 17–24. Tucson: University of Arizona Press.
Froeschlé C and Morbidelli A (1994) The secular resonances in the Solar System. In: Milani A, di Martino M, and Cellino A (eds.) *Asteroids, Comets, Meteors 1993, Vol. 160: IAU Symposium*, p. 189. Dordrecht: Kluwer Academic Publishers.
Fu R and Weiss BP (2011) Magnetic fields on asteroid 4 Vesta recorded by the Millbillillie eucrite. *EPSC-DPS Joint Meeting 2011*, p. 1646.
Fujiwara A, Kawaguchi J, Yeomans DK, et al. (2006) The rubble-pile asteroid Itokawa as observed by Hayabusa. *Science* 312: 1330–1334.
Fulvio D, Brunetto R, Vernazza P, and Strazzulla G (2012) Space weathering of Vesta and V-type asteroids: New irradiation experiments on HED meteorites. *Astronomy and Astrophysics* 537: L11.
Gaffey MJ (2010) Space weathering and the interpretation of asteroid reflectance spectra. *Icarus* 209: 564–574.
Gaffey MJ, Burbine TH, Piatek JL, et al. (1993) Mineralogical variations within the S-type asteroid class. *Icarus* 106: 573.
Gaffey MJ, Cloutis EA, Kelley MS, and Reed KL (2002) Mineralogy of asteroids. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 183–204. Tucson: University of Arizona Press.
Gehrels T, Drummond JD, and Levenson NA (1987) The absence of satellites of asteroids. *Icarus* 70: 257–263.
Geissler P, Petit J-M, Durda DD, et al. (1996) Erosion and ejecta reaccretion on 243 Ida and its moon. *Icarus* 120: 140–157.
Gilbert AM and Wiegert PA (2010) Updated results of a search for main-belt comets using the Canada-France-Hawaii Telescope Legacy Survey. *Icarus* 210: 998–999.
Gladman BJ, Davis DR, Neese C, et al. (2009) On the asteroid belt's orbital and size distribution. *Icarus* 202: 104–118.
Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.
Gradie JC, Chapman CR, and Tedesco EF (1989) Distribution of taxonomic classes and the compositional structure of the asteroid belt. In: Binzel RP, Gehrels T, and Matthews MS (eds.) *Asteroids II*, pp. 316–335. Tucson: University of Arizona Press.
Greenberg R (2008) Eros' Rahe Dorsum: Implications for internal structure. *Meteoritics and Planetary Science* 43: 435–449.
Greenberg R, Nolan MC, Bottke WF Jr., et al. (1994) Collisional history of Gaspra. *Icarus* 107: 84.
Groussin O, A'Hearn MF, Li J-Y, et al. (2007) Surface temperature of the nucleus of Comet 9P/Tempel 1. *Icarus* 187: 16–25.
Groussin O, Sunshine JM, Feaga LM, et al. (2013) The temperature, thermal inertia, roughness and color of the nuclei of Comets 103P/Hartley 2 and 9P/Tempel 1. *Icarus* 222: 580–594.
Grün E and Jessberger EK (1990) Dust. In: Huebner WF (ed.) *Physics and Chemistry of Comets*, pp. 113–175. Berlin: Springer-Verlag.
Halliday AN (2013) The origins of volatiles in the terrestrial planets. *Geochimica et Cosmochimica Acta* 105: 146–171.
Hapke B (1973) Darkening of silicate rock powders by solar wind sputtering. *Moon* 7: 342–355.
Hapke B (2001) Space weathering from Mercury to the asteroid belt. *Journal of Geophysical Research* 106: 10039–10074.
Harmon JK, Nolan MC, Giorgini JD, and Howell ES (2010) Radar observations of 8P/Tuttle: A contact-binary comet. *Icarus* 207: 499–502.
Harmon JK, Nolan MC, Howell ES, Giorgini JD, and Taylor PA (2011) Radar observations of comet 103P/Hartley 2. *The Astrophysical Journal Letters* 734: L2.
Harmon JK, Nolan MC, Margot J-L, et al. (2006) Radar observations of Comet P/2005 JQ5 (Catalina). *Icarus* 184: 285–288.
Harmon JK, Nolan MC, Ostro SJ, and Campbell DB (2004) Radar studies of comet nuclei and grain comae. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 265–279. Tucson: University of Arizona Press.
Hartogh P, Lis DC, Bockelée-Morvan D, et al. (2011) Ocean-like water in the Jupiter-family comet 103P/Hartley 2. *Nature* 478: 218–220.
Hirata N, Barnouin-Jha OS, Honda C, et al. (2009) A survey of possible impact structures on 25143 Itokawa. *Icarus* 200: 486–502.
Hiroi T, Abe M, Kitazato K, et al. (2006) Developing space weathering on the asteroid 25143 Itokawa. *Nature* 443: 56–58.
Hiroi T, Zolensky ME, and Pieters CM (2001) The Tagish Lake meteorite: A possible sample from a D-type asteroid. *Science* 293: 2234–2236.
Holsapple K, Giblin I, Housen K, Nakamura A, and Ryan E (2002) Asteroid impacts: Laboratory experiments and scaling laws. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 443–462. Tucson: University of Arizona Press.
Holsapple KA and Housen KR (2007) A crater and its ejecta: An interpretation of Deep Impact. *Icarus* 187: 345–356.
Horner J, Evans NW, and Bailey ME (2004) Simulations of the population of Centaurs. I: The bulk statistics. *Monthly Notices of the Royal Astronomical Society* 354: 798–810.
Housen KR, Holsapple KA, and Voss ME (1999) Compaction as the origin of the unusual craters on the asteroid Mathilde. *Nature* 402: 155–157.
Housen KR, Wilkening LL, Chapman CR, and Greenberg R (1979a) Asteroidal regoliths. *Icarus* 39: 317–351.
Housen KR, Wilkening LL, Chapman CR, and Greenberg RJ (1979b) Regolith development and evolution on asteroids and the moon. In: Gehrels T (ed.) *Asteroids*, pp. 601–627. Tucson: University of Arizona Press.
Howell ES, Nolan MC, Harmon JK, et al. (2007) Radar and radio observations of the fragmented comet 73P/Schwassmann-Wachmann3. *AAS/Division for Planetary Sciences Meeting Abstracts* 39: 486.
Hsieh HH (2009) The Hawaii trails project: Comet-hunting in the main asteroid belt. *Astronomy and Astrophysics* 505: 1297–1310.
Hsieh HH, Jewitt DC, and Fernández YR (2004) The strange case of 133P/Elst-Pizarro: A comet among the asteroids. *Astronomical Journal* 127: 2997–3017.
Hsieh HH, Jewitt D, Lacerda P, Lowry SC, and Snodgrass C (2010) The return of activity in main-belt comet 133P/Elst-Pizarro. *Monthly Notices of the Royal Astronomical Society* 403: 363–377.
Hsieh HH, Pohlen M, and Matulonis A (2013) Comet 133P/Elst-Pizarro. *Central Bureau Electronic Telegrams* 3564: 1.
Hudson RS and Ostro SJ (1994) Shape of asteroid 4769 Castalia (1989 PB) from inversion of radar images. *Science* 263: 940–943.
Hughes ALH and Armitage PJ (2010) Particle transport in evolving protoplanetary disks: implications for results from Stardust. *Astrophysical Journal* 719: 1633–1653.
IAU (2006) Resolution B5: Definition of a planet in the Solar System. www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf. Retrieved June 1, 2013.
Irvine WM, Bergin EA, Dickens JE, et al. (1998) Chemical processing in the coma as the source of cometary HNC. *Nature* 393: 547–550.
Ishiguro M, Ham J-B, Tholen DJ, et al. (2011) Search for the comet activity of 107P/(4015) Wilson-Harrington during 2009/2010 apparition. *Astrophysical Journal* 726: 101.
Ishiguro M, Hiroi T, Tholen DJ, et al. (2007) Global mapping of the degree of space weathering on asteroid 25143 Itokawa by Hayabusa/AMICA observations. *Meteoritics and Planetary Science* 42: 1791–1800.
Ivanov BA, Neukum G, Bottke WF Jr., and Hartmann WK (2002) The comparison of size-frequency distributions of impact craters and asteroids and the planetary cratering rate. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 89–101. Tucson: University of Arizona Press.
Izenberg NR, Murchie SL, Bell JF III, et al. (2003) Spectral properties and geologic processes on Eros from combined NEAR NIS and MSI data sets. *Meteoritics and Planetary Science* 38: 1053–1077.
Jacobson SA and Scheeres DJ (2011) Dynamics of rotationally fissioned asteroids: Source of observed small asteroid systems. *Icarus* 214: 161–178.

Jaumann R, Williams DA, Buczkowski DL, et al. (2012) Vesta's shape and morphology. *Science* 336: 687.

Jedicke R, Larsen J, and Spahr T (2002) Observational selection effects in asteroid surveys. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 71–87. Tucson: University of Arizona Press.

Jewitt D and Haghighipour N (2007) Irregular satellites of the planets: Products of capture in the early Solar System. *Annual Review of Astronomy and Astrophysics* 45: 261–295.

Johansen A, Oishi JS, Mac Low M-M, et al. (2007) Rapid planetesimal formation in turbulent circumstellar disks. *Nature* 448: 1022–1025.

Kavelaars JJ, Mousis O, Petit JM, et al. (2011) On the formation location of Uranus and Neptune as constrained by dynamical and chemical models of comets. *Astrophysical Journal* 734: L30.

Keil K (2002) Geological history of asteroid 4 Vesta: The "smallest terrestrial planet" In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 573–584. Tucson: University of Arizona Press.

Keller HU, Arpigny C, Barbieri C, et al. (1986) First Halley multicolour camera imaging results from Giotto. *Nature* 321: 320–326.

Keller LP and McKay DS (1993) Discovery of vapor deposits in the lunar regolith. *Science* 261: 1305–1307.

Keller HU, Thomas N, Curt W, et al. (1992) Giotto Halley Multicolor Camera Images V1.0, GIO-C-HMC-3-RDR-HALLEY-V1.0, NASA Planetary Data System.

Kelley MS and Wooden DH (2009) The composition of dust in Jupiter-family comets inferred from infrared spectroscopy. *Planetary and Space Science* 57: 1133–1145.

Kemper F, Vriend WJ, and Tielens AGGM (2004) The absence of crystalline silicates in the diffuse interstellar medium. *Astrophysical Journal* 609: 826–837.

Knight MM, Schleicher DG, Farnham TL, Schwieterman EW, and Christensen SR (2012) A quarter-century of observations of comet 10P/Tempel 2 at Lowell Observatory: Continued spin-down, coma morphology, production rates, and numerical modeling. *Astronomical Journal* 144: 153.

Krohn K, Jaumann R, and Otto K (in press) Mass movement on Vesta at steep scarps and crater rims. *Icarus*.

Küppers M, Moissl R, Vincent J-B, et al. (2012) Boulders on Lutetia. *Planetary and Space Science* 66: 71–78.

Küppers M, O'Rourke L, Bockelée-Morvan D, et al. (2014) Localized sources of water vapour on the dwarf planet (1) Ceres. *Nature* 505: 525–527.

Langland-Shula LE and Smith GH (2011) Comet classification with new methods for gas and dust spectroscopy. *Icarus* 213: 280–322.

Lauretta DS and McSween HY (2006) *Meteorites and the Early Solar System II*. Tucson: University of Arizona Press.

Lawler SM, Beichman CA, Bryden G, et al. (2009) Explorations beyond the snow line: Spitzer/IRS spectra of debris disks around solar-type stars. *Astrophysical Journal* 705: 89–111.

Lebofsky LA (1978) Asteroid 1 Ceres – Evidence for water of hydration. *Monthly Notices of the Royal Astronomical Society* 182: 17P–21P.

Lebofsky LA (1980) Infrared reflectance spectra of asteroids: A search for water of hydration. *Astronomical Journal* 85: 573–585.

Levison HF (1996) Comet taxonomy. In: Rettig T and Hahn JM (eds.) *Completing the Inventory of the Solar System, Vol. 107: Astronomical Society of the Pacific Conference Series*, pp. 173–191. San Francisco: Astronomical Society of the Pacific.

Li J-Y, A'Hearn MF, Belton MJS, et al. (2013) Photometry of the nucleus of Comet 9P/Tempel 1 from Stardust-NExT flyby and the implications. *Icarus* 222: 467–476.

Licandro J, Campins H, Kelley M, et al. (2011) Cybele: Detection of small silicate grains, water-ice, and organics. *Astronomy and Astrophysics* 525: A34.

Lis DC, Biver N, Bockelee-Morvan D, et al. (2013) A Herschel study of D/H in water in the Jupiter-family comet 45P/Honda-Mrkos-Pajdusakova and prospects for D/H measurements with CCAT. *Astrophysical Journal* 774: L3.

Lisse CM, Beichman CA, Bryden G, and Wyatt MC (2007) On the nature of the dust in the debris disk around HD 69830. *Astrophysical Journal* 658: 584–592.

Loeffler MJ, Baragiola RA, and Murayama M (2008) Laboratory simulations of redeposition of impact ejecta on mineral surfaces. *Icarus* 196: 285–292.

Loeffler MJ, Dukes CA, and Baragiola RA (2009) Irradiation of olivine by 4 keV He^{+}: Simulation of space weathering by the solar wind. *Journal of Geophysical Research: Planets* 114: E03003.

Longobardo A, Palomba E, Girasole M, et al. (2011) Study of iron nanophases in ordinary chondrites by means of near field microscopy. *Memorie della Societa Astronomica Italiana Supplementi* 16: 141.

Magri C, Howell ES, Nolan MC, et al. (2011) Radar and photometric observations and shape modeling of contact binary near-Earth Asteroid (8567) 1996 HW1. *Icarus* 214: 210–227.

Magri C, Nolan MC, Ostro SJ, and Giorgini JD (2007) A radar survey of main-belt asteroids: Arecibo observations of 55 objects during 1999–2003. *Icarus* 186: 126–151.

Marchi S, Barbieri C, Küppers M, et al. (2010a) The cratering history of asteroid (2867) Steins. *Planetary and Space Science* 58: 1116–1123.

Marchi S, Brunetto R, Magrin S, Lazzarin M, and Gandolfi D (2005) Space weathering of near-Earth and main belt silicate-rich asteroids: Observations and ion irradiation experiments. *Astronomy and Astrophysics* 443: 769–775.

Marchi S, De Sanctis MC, Lazzarin M, and Magrin S (2010b) On the puzzle of space weathering alteration of basaltic asteroids. *The Astrophysical Journal Letters* 721: L172–L176.

Marchi S, Massironi M, Cremonese G, et al. (2011) The effects of the target material properties and layering on the crater chronology: The case of Raditladi and Rachmaninoff basins on Mercury. *Planetary and Space Science* 59: 1968–1980.

Marchi S, Massironi M, Vincent J-B, et al. (2012a) The cratering history of asteroid (21) Lutetia. *Planetary and Space Science* 66: 87–95.

Marchi S, McSween HY, O'Brien DP, et al. (2012b) The violent collisional history of asteroid 4 Vesta. *Science* 336: 690.

Margot JL, Nolan MC, Benner LAM, et al. (2002) Binary asteroids in the near-Earth object population. *Science* 296: 1445–1448.

Margot JL, Nolan MC, Negron V, et al. (2003) 1937 UB (Hermes). *IAU Circular* 8227: 2.

Marsden BG, Sekanina Z, and Yeomans DK (1973) Comets and nongravitational forces. V. *Astronomical Journal* 78: 211.

Marty B (2012) The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth and Planetary Science Letters* 313: 56–66.

Marzari F and Scholl H (2002) On the instability of Jupiter's Trojans. *Icarus* 159: 328–338.

Masiero JR, Mainzer AK, Grav T, et al. (2011) Main belt asteroids with WISE/NEOWISE. I. Preliminary albedos and diameters. *Astrophysical Journal* 741: 68.

Massironi M, Marchi S, Pajola M, et al. (2012) Geological map and stratigraphy of asteroid 21 Lutetia. *Planetary and Space Science* 66: 125–136.

Matsui T and Abe Y (1986) Impact-induced atmospheres and oceans on Earth and Venus. *Nature* 322: 526–528.

McCord TB, Adams JB, and Johnson TV (1970) Asteroid Vesta: Spectral reflectivity and compositional implications. *Science* 168: 1445–1447.

McFadden LA, Gaffey MJ, and McCord TB (1984) Mineralogical-petrological characterization of near-Earth asteroids. *Icarus* 59: 25–40.

McFadden LA, Wellnitz DD, Schnaubelt M, et al. (2001) Mineralogical interpretation of reflectance spectra of Eros from NEAR near-infrared spectrometer low phase flyby. *Meteoritics and Planetary Science* 36: 1711–1726.

Meech KJ, Buie MW, Samarasinha NH, Mueller BEA, and Belton MJS (1997) Observations of structures in the inner coma of Chiron with the HST Planetary Camera. *Astronomical Journal* 113: 844–862.

Meech KJ, Hainaut OR, and Marsden BG (2004) Comet nucleus size distributions from HST and Keck telescopes. *Icarus* 170: 463–491.

Meech KJ, Pittichová J, Yang B, et al. (2011) Deep Impact, Stardust-NExT and the behavior of Comet 9P/Tempel 1 from 1997 to 2010. *Icarus* 213: 323–344.

Meech KJ and Svoren J (2004) Using cometary activity to trace the physical and chemical evolution of cometary nuclei. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 317–335. Tucson: University of Arizona Press.

Melosh HJ (1989) *Impact Cratering: A Geologic Process*. New York: Oxford University Press.

Merline WJ, Weidenschilling SJ, Durda DD, et al. (2002) Asteroids do have satellites. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 289–312. Tucson: University of Arizona Press.

Michel P, O'Brien DP, Abe S, and Hirata N (2009) Itokawa's cratering record as observed by Hayabusa: Implications for its age and collisional history. *Icarus* 200: 503–513.

Michikami T, Nakamura AM, and Hirata N (2010) The shape distribution of boulders on Asteroid 25143 Itokawa: Comparison with fragments from impact experiments. *Icarus* 207: 277–284.

Michikami T, Nakamura AM, Hirata N, et al. (2008) Size-frequency statistics of boulders on global surface of asteroid 25143 Itokawa. *Earth, Planets and Space* 60: 13–20.

Miyamoto H, Yano H, Scheeres DJ, et al. (2007) Regolith migration and sorting on asteroid Itokawa. *Science* 316: 1011.

Morbidelli A (2010) A coherent and comprehensive model of the evolution of the outer Solar System. *Comptes Rendus Physique* 11: 651–659.

Morbidelli A and Brown ME (2004) The Kuiper belt and the primordial evolution of the Solar System. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 175–191. Tucson: University of Arizona Press.

Morbidelli A, Chambers J, Lunine JI, et al. (2000) Source regions and timescales for the delivery of water on Earth. *Meteoritics and Planetary Science* 35: 1309–1320.

Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.

Morbidelli A, Lunine JI, O'Brien DP, Raymond SN, and Walsh KJ (2012) Building terrestrial planets. *Annual Review of Earth and Planetary Sciences* 40: 251–275.

Moretti PF, Maras A, Palomba E, et al. (2005) Detection of nanostructured metal in meteorites: Implications for the reddening of asteroids. *The Astrophysical Journal Letters* 634: L117–L120.

Moretti PF, Maras A, Somma F, and Aloe P (2006) Simulation of low-velocity bombardment of asteroid surfaces: regolith formation and reddening of the spectra. *Astronomy and Astrophysics* 459: 245–248.
Mothé-Diniz T, Jasmin FL, Carvano JM, et al. (2010) Re-assessing the ordinary chondrites paradox. *Astronomy and Astrophysics* 514: A86.
Mottl M, Glazer B, Kaiser R, and Meech K (2007) Water and astrobiology. *Chemie der Erde* 67: 253–282.
Mousis O, Gautier D, Bockelee-Morvan D, Robert F, Dubrulle B, and Drouart A (2000) Constraints on the formation of comets from D/H ratios measured in H_2O and HCN. *Icarus* 148: 513.
Mueller BEA and Ferrin I (1996) Change in the rotational period of comet P/Tempel 2 between the 1988 and 1994 apparitions. *Icarus* 123: 463–477.
Mueller BEA, Samarasinha NH, and Fernandez YR (2008) Near-aphelion lightcurves of comet 2P/Encke in 2008. *AAS/Division for Planetary Sciences Meeting Abstracts* 40: 416.
Mumma MJ and Charnley SB (2011) The chemical composition of comets—Emerging taxonomies and natal heritage. *Annual Review of Astronomy and Astrophysics* 49: 471–524.
Mumma MJ, Disanti MA, dello Russo N, et al. (2003) Remote infrared observations of parent volatiles in comets: A window on the early Solar System. *Advances in Space Research* 31: 2563–2575.
Murchie S, Robinson M, Clark B, et al. (2002) Color variations on Eros from NEAR multispectral imaging. *Icarus* 155: 145–168.
Nakamura T, Noguchi T, Tanaka M, et al. (2011) Itokawa dust particles: A direct link between S-type asteroids and ordinary chondrites. *Science* 333: 1113.
Nakamura T, Noguchi T, Tsuchiyama A, et al. (2008) Chondrulelike objects in short-period comet 81P/Wild 2. *Science* 321: 1664.
Nelson RM, Soderblom LA, and Hapke BW (2004) Are the circular, dark features on comet Borrelly's surface albedo variations or pits? *Icarus* 167: 37–44.
Nesvorný D, Bottke WF Jr., Vokrouhlický D, Chapman CR, and Rafkin S (2010) Do planetary encounters reset surfaces of near Earth asteroids? *Icarus* 209: 510–519.
Nesvorný D, Jedicke R, Whiteley RJ, and Ivezic Ž (2005) Evidence for asteroid space weathering from the Sloan Digital Sky Survey. *Icarus* 173: 132–152.
Newburn RL and Farnham TL (2008) Stardust Navcam Calibrated Images of 81P/WILD 2, SDU-C-NAVCAM-3-RDR-WILD2-V2.0, NASA Planetary Data System.
Noble SK, Keller LP, and Pieters CM (2010) Evidence of space weathering in regolith breccias II: Asteroidal regolith breccias. *Meteoritics and Planetary Science* 45: 2007–2015.
Noble SK, Pieters CM, and Keller LP (2007) An experimental approach to understanding the optical effects of space weathering. *Icarus* 192: 629–642.
Noguchi T, Kimura M, Hashimoto T, et al. (2014) Space weathered rims found on the surfaces of the Itokawa dust particles. *Meteoritics and Planetary Science* 49: 188–214.
Noguchi T, Nakamura T, Kimura M, et al. (2011) Incipient space weathering observed on the surface of Itokawa dust particles. *Science* 333: 1121.
Nolan MC, Hine AA, Howell ES, et al. (2005) Extreme diversity of near-Earth asteroid physical properties from Arecibo radar imaging. *American Astronomical Society Meeting Abstracts* 37: 1155.
Nolan MC, Magri C, Benner LAM, et al. (2012) The shape of OSIRIS-REx mission target 1999 RQ36 from radar and lightcurve data. *LPI Contributions* 1667: 6345.
O'Brien DP (2009) The Yarkovsky effect is not responsible for small crater depletion on Eros and Itokawa. *Icarus* 203: 112–118.
O'Brien DP and Greenberg R (2005) The collisional and dynamical evolution of the main-belt and NEA size distributions. *Icarus* 178: 179–212.
O'Brien DP, Greenberg R, and Richardson JE (2006) Craters on asteroids: Reconciling diverse impact records with a common impacting population. *Icarus* 183: 79–92.
Ogliore RC, Westphal AJ, Gainsforth Z, et al. (2009) Nebular mixing constrained by the Stardust samples. *Meteoritics and Planetary Science* 44: 1675–1681.
Ootsubo T, Usui F, Kawakita H, et al. (2010) Detection of parent H_2O and CO_2 molecules in the 2.5–5 μm spectrum of comet C/2007 N3 (Lulin) observed with AKARI. *The Astrophysical Journal Letters* 717: L66–L70.
Ostro SJ (1989) Radar observations of asteroids. In: Binzel RP, Gehrels T, and Matthews MS (eds.) *Asteroids II*, pp. 192–212. Tucson: University of Arizona Press.
Ostro SJ (1993) Planetary radar astronomy. *Reviews of Modern Physics* 65: 1235–1279.
Ostro SJ, Benner LAM, Magri C, et al. (2005) Radar observations of Itokawa in 2004 and improved shape estimation. *Meteoritics and Planetary Science* 40: 1563–1574.
Ostro SJ, Hudson RS, Benner LAM, et al. (2002) Asteroid radar astronomy. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 151–168. Tucson: University of Arizona Press.
Ostro SJ, Margot J-L, Benner LAM, et al. (2006) Radar imaging of binary near-Earth asteroid (66391) 1999 KW4. *Science* 314: 1276–1280.
Pieters CM, Ammannito E, Blewett DT, et al. (2012) Distinctive space weathering on Vesta from regolith mixing processes. *Nature* 491: 79–82.
Pieters CM, Binzel RP, Bogard D, et al. (2006) Asteroid meteorite links: The Vesta conundrum(s). In: Sylvio Ferraz M and Angel FJ (eds.) *Asteroids, Comets, Meteors, Vol. 229: IAU Symposium*, pp. 273–288. Cambridge: Cambridge University Press.
Pieters CM, Taylor LA, Noble SK, et al. (2000) Space weathering on airless bodies: Resolving a mystery with lunar samples. *Meteoritics and Planetary Science* 35: 1101–1107.
Pizzarello S, Cooper GW, and Flynn GJ (2006) The nature and distribution of the organic material in carbonaceous chondrites and interplanetary dust particles. In: Lauretta DS and McSween HY Jr. (eds.) *Meteorites and the Early Solar System II*, pp. 625–651. Tucson: University of Arizona Press.
Pravec P and Harris AW (2000) Fast and slow rotation of asteroids. *Icarus* 148: 12–20.
Pravec P, Scheirich P, Kušnirák P, et al. (2006) Photometric survey of binary near-Earth asteroids. *Icarus* 181: 63–93.
Prialnik D, Benkhoff J, and Podolak M (2004) Modeling the structure and activity of comet nuclei. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 359–387. Tucson: University of Arizona Press.
Prialnik D, Sarid G, Rosenberg ED, and Merk R (2008) Thermal and chemical evolution of comet nuclei and Kuiper belt objects. *Space Science Reviews* 138: 147–164.
Prockter L, Thomas P, Robinson M, et al. (2002) Surface expressions of structural features on Eros. *Icarus* 155: 75–93.
Protopapa S, Sunshine JM, Feaga LM, et al. (2014) Water ice and dust in the innermost coma of 103P/Hartley 2. *Icarus* 238: 191–204.
Raymond SN, Quinn T, and Lunine JI (2004) Making other earths: Dynamical simulations of terrestrial planet formation and water delivery. *Icarus* 168: 1–17.
Reach WT, Kelley MS, and Vaubaillon J (2013) Survey of cometary CO_2, CO, and particulate emissions using the Spitzer Space Telescope: Smog check for comets. ArXiv e-prints 1306.2381.
Reinhard R (1986) The Giotto encounter with comet Halley. *Nature* 321: 313–318.
Richardson DC (2002) Gravitational aggregates: Evidence and evolution. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 501–515. Tucson: University of Arizona Press.
Richardson JE, Melosh HJ, and Greenberg R (2004) Impact-induced seismic activity on asteroid 433 Eros: A surface modification process. *Science* 306: 1526–1529.
Richardson JE, Melosh HJ, Lisse CM, and Carcich B (2007) A ballistics analysis of the Deep Impact ejecta plume: Determining Comet Tempel 1's gravity, mass, and density. *Icarus* 190: 357–390.
Rivkin AS (2012) The fraction of hydrated C-complex asteroids in the asteroid belt from SDSS data. *Icarus* 221: 744–752.
Rivkin AS and Emery JP (2010) Detection of ice and organics on an asteroidal surface. *Nature* 464: 1322–1323.
Rivkin AS, Howell ES, Vilas F, and Lebofsky LA (2002) Hydrated minerals on asteroids: The astronomical record. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 235–253. Tucson: University of Arizona Press.
Robinson MS, Thomas PC, Veverka J, Murchie S, and Carcich B (2001) The nature of ponded deposits on Eros. *Nature* 413: 396–400.
Rodgers SD and Charnley SB (1998) HNC and HCN in Comets. *The Astrophysical Journal Letters* 501: L227.
Romanishin W, Tegler SC, and Consolmagno GJ (2010) Colors of inner disk classical Kuiper belt objects. *Astronomical Journal* 140: 29–33.
Russell CT, Raymond CA, Coradini A, et al. (2012) Dawn at Vesta: Testing the protoplanetary paradigm. *Science* 336: 684.
Saal AE, Hauri EH, van Orman JA, and Rutherford MJ (2013) Hydrogen isotopes in lunar volcanic glasses and melt inclusions reveal a carbonaceous chondrite heritage. *Science* 340: 1317–1320.
Sagdeev RZ, Szabo F, Avanesov GA, et al. (1986) Television observations of comet Halley from VEGA spacecraft. *Nature* 321: 262–266.
Sakamoto N, Seto Y, Itoh S, et al. (2007) Remnants of the early Solar System water enriched in heavy oxygen isotopes. *Science* 317: 231.
Salisbury JW, D'Aria DM, and Jarosewich E (1991) Mid-infrared (2.5–13.5 μm) reflectance spectra of powdered stony meteorites. *Icarus* 92: 280–297.
Samarasinha NH, Mueller BEA, A'Hearn MF, Farnham TL, and Gersch A (2011) Rotation of comet 103P/Hartley 2 from structures in the coma. *The Astrophysical Journal Letters* 734: L3.
Samarasinha NH, Mueller BEA, Belton MJS, and Jorda L (2004) Rotation of cometary nuclei. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 281–299. Tucson: University of Arizona Press.
Sasaki S, Nakamura K, Hamabe Y, Kurahashi E, and Hiroi T (2001) Production of iron nanoparticles by laser irradiation in a simulation of lunar-like space weathering. *Nature* 410: 555–557.
Scheeres DJ, Fahnestock EG, Ostro SJ, et al. (2006) Dynamical configuration of binary near-Earth asteroid (66391) 1999 KW4. *Science* 314: 1280–1283.
Schenk P, O'Brien DP, Marchi S, et al. (2012) The geologically recent giant impact basins at Vesta's South pole. *Science* 336: 694.

Schleicher DG, Millis RL, Osip DJ, and Birch PV (1991) Comet Levy (1990c) – Groundbased photometric results. *Icarus* 94: 511–523.

Schleicher DG and Woodney LM (2003) Analyses of dust coma morphology of Comet Hyakutake (1996 B2) near perigee: Outburst behavior, jet motion, source region locations, and nucleus pole orientation. *Icarus* 162: 190–213.

Schorghofer N (2008) The lifetime of ice on main belt asteroids. *Astrophysical Journal* 682: 697–705.

Schultz PH, Hermalyn B, and Veverka J (2013) The Deep Impact crater on 9P/Tempel-1 from Stardust-NExT. *Icarus* 222: 502–515.

Schulz R (2009) Rosetta—One comet rendezvous and two asteroid fly-bys. *Solar System Research* 43: 343–352.

Shepard MK, Clark BE, Ockert-Bell M, et al. (2010) A radar survey of M- and X-class asteroids II: Summary and synthesis. *Icarus* 208: 221–237.

Shepard MK, Taylor PA, Nolan MC, et al. (submitted for publication). A radar survey of M- and X-class asteroids III: Insights into their composition, hydration state, and structure. *Icarus*.

Sheppard SS (2012) The color differences of Kuiper belt objects in resonance with Neptune. *Astronomical Journal* 144: 169.

Shingareva TV, Basilevsky AT, Shashkina VP, et al. (2008) Morphological characteristics of the Phobos craters and grooves. *Lunar and Planetary Institute Science Conference Abstracts* 39: 2425.

Shu F, Najita J, Ostriker E, et al. (1994) Magnetocentrifugally driven flows from young stars and disks. 1: A generalized model. *Astrophysical Journal* 429: 781–796.

Shu FH, Najita JR, Shang H, and Li Z-Y (2000) X-winds theory and observations. In: Mannings V, Boss AP, Russell SS (eds.), *Protostars and Planets* IV, pp. 789–814. Tucson: University of Arizona Press.

Sierks H, Lamy P, Barbieri C, et al. (2011) Images of asteroid 21 Lutetia: A remnant planetesimal from the early Solar System. *Science* 334: 487.

Sitko ML, Lisse CM, Kelley MS, et al. (2011) Infrared spectroscopy of comet 73P/Schwassmann-Wachmann 3 using the Spitzer Space Telescope. *Astronomical Journal* 142: 80.

Stadermann FJ, Hoppe P, Floss C, et al. (2008) Stardust in Stardust – The C, N, and O isotopic compositions of Wild 2 cometary matter in Al foil impacts. *Meteoritics and Planetary Science* 43: 299–313.

Strazzulla G, Dotto E, Binzel R, et al. (2005) Spectral alteration of the Meteorite Epinal (H5) induced by heavy ion irradiation: a simulation of space weathering effects on near-Earth asteroids. *Icarus* 174: 31–35.

Sullivan R, Greeley R, Pappalardo R, et al. (1996) Geology of 243 Ida. *Icarus* 120: 119–139.

Sunshine JM, A'Hearn MF, Groussin O, et al. (2006) Exposed water ice deposits on the surface of comet 9P/Tempel 1. *Science* 311: 1453–1455.

Sunshine JM, Feaga LM, Groussin O, et al. (2011) Water ice on comet 103P/Hartley 2. In: *EPSC-DPS Joint Meeting 2011*, p. 1345.

Sunshine JM, Groussin O, Schultz PH, et al. (2007) The distribution of water ice in the interior of Comet Tempel 1. *Icarus* 190: 284–294.

Taylor PA, Howell ES, Nolan MC, and Thane AA (2012) The shape and spin distributions of near-Earth asteroids observed with the Arecibo radar system. *LPI Contributions* 1667: 6340.

Taylor PA, Howell ES, Nolan MC, et al. (2013) Physical characterization of binary near-Earth asteroid (153958) 2002 AM31. *AAS/Division for Planetary Sciences Meeting Abstracts* 45: 208.08.

Tholen DJ and Barucci MA (1989) Asteroid taxonomy. In: Binzel RP, Gehrels T, and Matthews MS (eds.) *Asteroids II*, pp. 298–315. Tucson: University of Arizona Press.

Thomas P, A'Hearn M, Belton MJS, et al. (2013a) The nucleus of Comet 9P/Tempel 1: Shape and geology from two flybys. *Icarus* 222: 453–466.

Thomas PC, A'Hearn MF, Veverka J, et al. (2013b) Shape, density, and geology of the nucleus of Comet 103P/Hartley 2. *Icarus* 222: 550–558.

Thomas N, Barbieri C, Keller HU, et al. (2012) The geomorphology of (21) Lutetia: Results from the OSIRIS imaging system onboard ESA's Rosetta spacecraft. *Planetary and Space Science* 66: 96–124.

Thomas PC, Binzel RP, Gaffey MJ, et al. (1997a) Impact excavation on asteroid 4 Vesta: Hubble Space Telescope results. *Science* 277: 1492–1495.

Thomas PC, Binzel RP, Gaffey MJ, et al. (1997b) Vesta: Spin pole, size, and shape from HST images. *Icarus* 128: 88–94.

Thomas PC and Prockter LM (2010) Tectonics of small bodies. In: Watters TR and Schultz RA (eds.) *Planetary Tectonics*, pp. 391–423. Cambridge University Press.

Thomas PC, Prockter L, Robinson M, Joseph J, and Veverka J (2002) Global structure of asteroid 433 Eros. *Geophysical Research Letters* 29: 1408.

Thomas PC, Veverka J, Belton MJS, et al. (2007) The shape, topography, and geology of Tempel 1 from Deep Impact observations. *Icarus* 187: 4–15.

Thomas P, Veverka J, Bloom A, and Duxbury T (1979) Grooves on PHOBOS – Their distribution, morphology and possible origin. *Journal of Geophysical Research* 84: 8457–8477.

Thomas PC, Veverka J, Robinson MS, and Murchie S (2001) Shoemaker crater as the source of most ejecta blocks on the asteroid 433 Eros. *Nature* 413: 394–396.

Thomas PC, Veverka J, Simonelli D, et al. (1994) The shape of Gaspra. *Icarus* 107: 23.

Thomas PC, Veverka J, Sullivan R, et al. (2000) Phobos: Regolith and ejection blocks investigated with the Mars Orbiter Camera images. *Journal of Geophysical Research* 105: 15091–15106.

Tiscareno MS and Malhotra R (2003) The dynamics of known Centaurs. *Astronomical Journal* 126: 3122–3131.

Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Nature* 435: 459–461.

Turrini D, Marzari F, and Tosi F (2009) A new perspective on the irregular satellites of Saturn: II. Dynamical and physical origin. *Monthly Notices of the Royal Astronomical Society* 392: 455–474.

Udry S and Santos NC (2007) Statistical properties of exoplanets. *Annual Review of Astronomy and Astrophysics* 45: 397–439.

Van Dishoeck EF, Bergin EA, Lis DC, and Lunine JI (2014) Water: From clouds to planets. In: Beuther H, Klessen R, Dullemond C, and Henning T (eds.) *Protostars and Planets IV*. Tucson: University of Arizona Press.

Vernazza P, Brunetto R, Strazzulla G, et al. (2006) Asteroid colors: A novel tool for magnetic field detection? The case of Vesta. *Astronomy and Astrophysics* 451: L43–L46.

Veverka J, Belton M, Klaasen K, and Chapman C (1994) Galileo's encounter with 951 Gaspra: Overview. *Icarus* 107: 2–17.

Veverka J, Robinson M, Thomas P, et al. (2000) NEAR at Eros: Imaging and spectral results. *Science* 289: 2088–2097.

Veverka J, Thomas PC, Bell JF III, et al. (1999) Imaging of asteroid 433 Eros during NEAR's flyby reconnaissance. *Science* 285: 562–564.

Veverka J, Thomas P, Johnson TV, et al. (1986) The physical characteristics of satellite surfaces. In: *Satellites*, pp. 342–402. Tucson: University of Arizona Press.

Vilas F (1994) A cheaper, faster, better way to detect water of hydration on Solar System bodies. *Icarus* 111: 456–467.

Vilas F and Gaffey MJ (1989) Phyllosilicate absorption features in main-belt and outer-belt asteroid reflectance spectra. *Science* 246: 790–792.

Vincent J-B, Besse S, Marchi S, et al. (2012a) Physical properties of craters on asteroid (21) Lutetia. *Planetary and Space Science* 66: 79–86.

Vincent J-B, Hoffman M, Nathues A, et al. (2012b) Crater depth-to-diameter ratio and surface properties of (4) Vesta. *Lunar and Planetary Institute Science Conference Abstracts* 43: 1415.

Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475: 206–209.

Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2012) Populating the asteroid belt from two parent source regions due to the migration of giant planets—"The Grand Tack" *Meteoritics and Planetary Science* 47: 1941–1947.

Warner BD, Harris AW, and Pravec P (2009) The asteroid lightcurve database. *Icarus* 202: 134–146.

Weissman PR, A'Hearn MF, Rickman H, and McFadden LA (1989) Evolution of comets into asteroids. In: Binzel RP, Gehrels T, and Matthews MS (eds.) *Asteroids II*, pp. 880–920. Tucson: University of Arizona Press.

Weissman PR, Bottke WF Jr., and Levison HF (2002) Evolution of comets into asteroids. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 669–686. Tucson: University of Arizona Press.

Whipple FL (1950) A comet model. I. The acceleration of Comet Encke. *Astrophysical Journal* 111: 375–394.

Williams DA, Denevi BW, Mittlefehldt DW, et al. (in press) The geology of the Marcia quadrangle of asteroid Vesta: Assessing the effects of large, young craters. *Icarus*.

Wooden DH (2008) Cometary refractory grains: interstellar and nebular sources. *Space Science Reviews* 138: 75–108.

Yamada M, Sasaki S, Fujiwara A, et al. (1999) Simulation of space weathering by nanosecond pulse laser heating and proton implantation: Difference of olivine and pyroxene samples. *Lunar and Planetary Institute Science Conference Abstracts* 30: 1566.

Yingst RA, Mest SC, Berman DC, et al. (submitted for publication) Geologic mapping of Vesta. *Icarus*.

Yurimoto H, Abe K, Abe M, et al. (2011) Oxygen isotopic compositions of asteroidal materials returned from Itokawa by the Hayabusa mission. *Science* 333: 1116.

Zolensky ME, Zega TJ, Yano H, et al. (2006) Mineralogy and petrology of comet 81P/Wild 2 nucleus samples. *Science* 314: 1735–1738.

Zuber MT, McSween HY, Binzel RP, et al. (2011) Origin, internal structure and evolution of 4 Vesta. *Space Science Reviews* 163: 77–93.

10.16 Giant Planets

T Guillot, Observatoire de la Côte d'Azur, CNRS, Université de Nice-Sophia Antipolis, Nice, France
D Gautier, LESIA, Observatoire de Paris, Meudon, France

© 2015 Elsevier B.V. All rights reserved.

10.16.1	**Introduction**	529
10.16.2	**Observations and Global Properties**	530
10.16.2.1	Visual Appearances	530
10.16.2.2	Gravity Fields	530
10.16.2.3	Magnetic Fields	531
10.16.2.4	Atmospheric Dynamics: Winds and Weather	532
10.16.2.5	Energy Balance and Atmospheric Temperature Profiles	533
10.16.2.6	Atmospheric Compositions	534
10.16.2.6.1	Hydrogen and helium	534
10.16.2.6.2	Heavy elements	534
10.16.2.7	Isotopic Ratios	537
10.16.2.8	Moons and Rings	538
10.16.2.9	Seismology	538
10.16.2.10	Exoplanets	538
10.16.3	**The Calculation of Interior and Evolution Models**	539
10.16.3.1	Basic Equations	539
10.16.3.2	High-Pressure Physics and Equations of State	540
10.16.3.2.1	Hydrogen	540
10.16.3.2.2	Other elements and mixtures	541
10.16.3.3	Heat Transport	542
10.16.3.4	The Contraction and Cooling Histories of the Giant Planets	542
10.16.3.5	Mass–Radius Relation	544
10.16.3.6	Rotation and the Figures of Planets	545
10.16.4	**Interior Structures and Evolutions**	546
10.16.4.1	Jupiter and Saturn	546
10.16.4.2	Uranus and Neptune	547
10.16.4.3	Irradiated Giant Planets	548
10.16.4.3.1	Interior structure and dynamics	548
10.16.4.3.2	Thermal evolution and inferred compositions	549
10.16.5	**Implications for Planetary Formation Models**	551
10.16.6	**Future Prospects**	552
Acknowledgments		553
References		553

10.16.1 Introduction

In our solar system, four planets stand out for their sheer mass and size. Jupiter, Saturn, Uranus, and Neptune indeed qualify as the 'giant planets' because they are larger than any terrestrial planet and much more massive than all other objects in the solar system, except the Sun, put together (Figure 1). Because of their gravitational might, they have played a key role in the formation of the solar system, tossing around many objects in the system, preventing the formation of a planet in what is now the asteroid belt, and directly leading to the formation of the Kuiper Belt and Oort Cloud. They also retain some of the gas (in particular hydrogen and helium) that was present when the Sun and its planets formed and are thus key witnesses in the search for our origins.

Because of a massive envelope mostly made of hydrogen–helium, these planets are *fluid*, with no solid or liquid surface. In terms of structure and composition, they lie in between stars (gaseous and mostly made of hydrogen and helium) and smaller terrestrial planets (solid and liquid and mostly made of heavy elements), with Jupiter and Saturn being closer to the former and Uranus and Neptune to the latter (see Figure 1).

The discovery of many extrasolar planets of masses from a few thousands down to a few Earth masses and the possibility to characterize them by the measurement of their mass and size prompt a more general definition of the giant planets. For this chapter, we will adopt the following: "a giant planet is a planet mostly made of hydrogen and helium and too light to ignite deuterium fusion." This is purposely relatively vague – depending on whether the inventory is performed by mass or by atom or molecule, Uranus and Neptune may be included or left out of the category. Note that Uranus and Neptune are indeed relatively different in structure than Jupiter and Saturn and are generally referred to as 'ice giants,' due to an interior

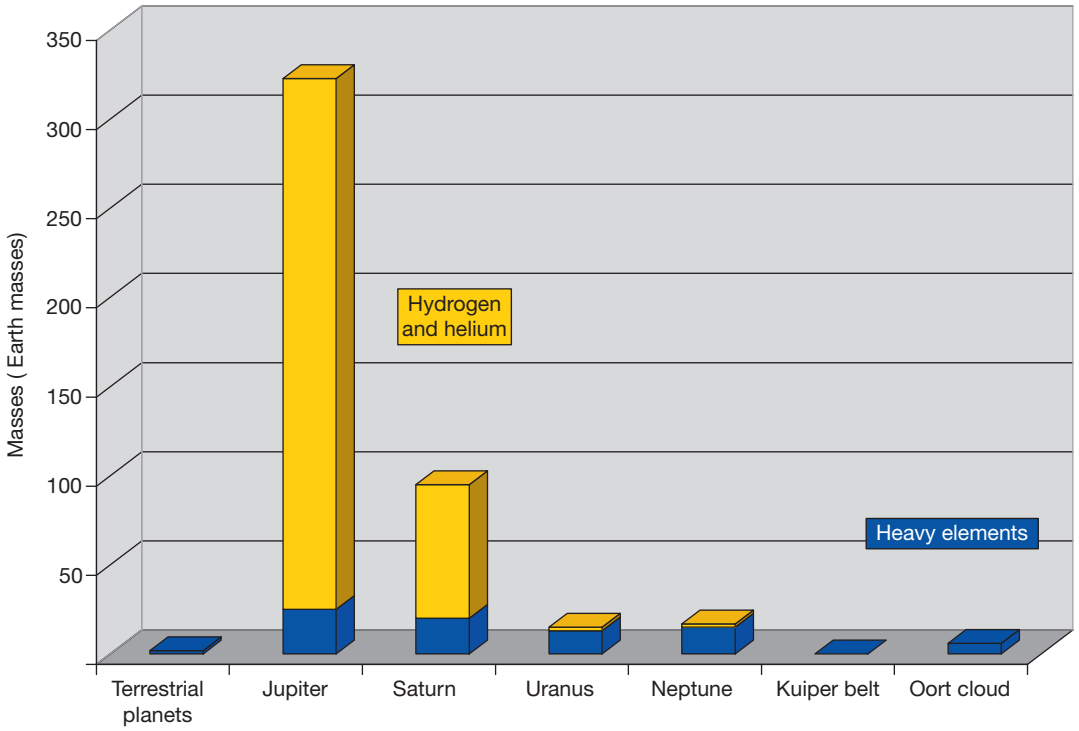


Figure 1 An inventory of hydrogen and helium and all other elements ('heavy elements') in the solar system excluding the Sun (the Sun has a total mass of 332 960 $M_\oplus$, including about 5000 $M_\oplus$ in heavy elements, 1 $M_\oplus$ being the mass of the Earth). The precise amount of heavy elements in Jupiter (10–40 $M_\oplus$) and Saturn (20–30 $M_\oplus$) is uncertain (see **Section 10.16.4.1**).

structure that is consistent with the presence of mostly 'ices' (a mixture formed from the condensation in the protoplanetary disk of low-refractivity materials such as H_2O, CH_4, and NH_3 and brought to the high-pressure conditions of planetary interiors – see in the succeeding text).

Globally, this definition encompasses a class of objects that have similar properties (in particular, a low viscosity and a non-negligible compressibility) and inherited part of their material directly from the same reservoir as their parent star. These objects can thus be largely studied with the same tools, and their formation is linked to that of their parent star and the fate of the circumstellar gaseous disk present around the young star.

We will hereafter present some of the key data concerning the giant planets in the solar system and outside. We will then present the theoretical basis for the study of their structure and evolution. On this basis, the constraints on their composition will be discussed and analyzed in terms of consequences for planet formation models.

10.16.2 Observations and Global Properties

10.16.2.1 Visual Appearances

In spite of its smallness, the sample of four giant planets in our solar system exhibits a large variety of appearances, shapes, colors, variability, etc. As shown in Figure 2, all four giant planets are flattened by rotation and exhibit a more or less clear zonal wind pattern, but the color of their visible atmosphere is very different (this is due mostly to minor species in the high planetary atmosphere), their clouds have different compositions (ammonia for Jupiter and Saturn and methane for Uranus and Neptune) and depths, and their global meteorology (number of vortexes, long-lived anticyclones such as Jupiter's Great Red Spot, the presence of planetary-scale storms, and convective activity) is different from one planet to the next.

We can presently only wonder about what is in store for us with extrasolar giant planets since we cannot image and resolve them. But with orbital distances from as close as 0.01 to 100 AU and more, a variety of masses, sizes, and parent stars, we should expect to be surprised!

10.16.2.2 Gravity Fields

The mass of our giant planets can be obtained with great accuracy from the observation of the motions of their natural satellites: 317.834, 95.161, 14.538, and 17.148 times the mass of the Earth ($1\ M_\oplus = 5.97369 \times 10^{27}$ g) for Jupiter, Saturn, Uranus, and Neptune, respectively. More precise measurements of their gravity field can be obtained through the analysis of the trajectories of a spacecraft during flyby, especially when they come close to the planet and preferably in a near-polar orbit. The gravitational field thus measured departs from a purely spherical function due to the planets' rapid rotation. The measurements are generally expressed by expanding the components of the gravity field in Legendre polynomials P_i of progressively higher orders:

Figure 2 Photomontage from images of Voyager 2 (Jupiter, Uranus, and Neptune) and Cassini (Saturn). The planets are shown to scale, with their respective axial inclinations.

$$V_{\text{ext}}(r,\theta) = -\frac{GM}{r}\left\{1 - \sum_{i=1}^{\infty}\left(\frac{R_{\text{eq}}}{r}\right)^{i} J_i P_i(\cos\theta)\right\} \qquad [1]$$

where $V_{\text{ext}}(r, \theta)$ is the gravity field evaluated outside the planet at a distance r and colatitude θ, R_{eq} is the equatorial radius, and J_i are the gravitational moments. Because the giant planets are very close to hydrostatic equilibrium, the coefficients of even order are the only ones that are not negligible. We will see how these gravitational moments, as listed in Table 1, help us constrain the planets' interior density profiles.

Table 1 also indicates the radii obtained with the greatest accuracy by radio-occultation experiments. An important consequence obtained is the fact that these planets have low densities, from 0.688 g cm^{-3} for Saturn to 1.64 g cm^{-3} for Neptune, to be compared with densities of 3.9–5.5 g cm^{-3} for the terrestrial planets in the solar system. Considering the compression that strongly increases with mass, one is led naturally to the conclusion that these planets contain an important proportion of light materials including hydrogen and helium. It also implies that Uranus and Neptune, which are less massive, must contain a relatively larger proportion of heavy elements than Jupiter and Saturn. This may lead to a subclassification between the hydrogen–helium giant planets Jupiter and Saturn and the 'ice giants' or 'subgiants' Uranus and Neptune.

The planets are also relatively fast rotators, with periods of ~10 h for Jupiter and Saturn and ~17 h for Uranus and Neptune. The fact that this fast rotation visibly affects the figure (shape) of these planets is seen by the significant difference between the polar and equatorial radii. It also leads to gravitational moments that differ significantly from a null value. However, it is important to stress that there is no unique rotation frame for these fluid planets: Atmospheric zonal winds imply that different latitudes rotate at different velocities (see Section 10.16.2.4), and the magnetic field provides another rotation period. Because the latter is tied to the deeper levels of the planet, it is believed to be more relevant when interpreting the gravitational moments. The rotation periods listed in Table 1 hence correspond to that of the magnetic field. The case of Saturn is complex and to be discussed in the next section.

10.16.2.3 Magnetic Fields

As the Earth, the Sun, and Mercury, our four giant planets possess their own magnetic fields. These magnetic fields **B** may be expressed in form of a development in spherical harmonics of the scalar potential W, such that $\mathbf{B} = -\nabla W$:

$$W = a\sum_{n=1}^{\infty}\left(\frac{a}{r}\right)^{n+1}\sum_{m=0}^{n}\left\{g_n^m\cos(m\phi) + h_n^m\sin(m\phi)\right\}P_n^m(\cos\theta) \qquad [2]$$

where r is the distance to the planet's center, a its radius, θ the colatitude, ϕ the longitude, and P_n^m the associated Legendre polynomials. The coefficients g_n^m and h_n^m are the magnetic moments that characterize the field. They are expressed in magnetic field units.

One can show that the first coefficients of relation [2] (for $n = 0$ and 1) correspond to the potential of a magnetic dipole such that $W = \mathbf{M}\cdot\mathbf{r}/r^3$ of moment:

$$M = a^3\left\{\left(g_1^0\right)^2 + \left(g_1^1\right)^2 + \left(h_1^1\right)^2\right\}^{1/2} \qquad [3]$$

As shown by the Voyager 2 measurements, Jupiter and Saturn have magnetic fields of essentially dipolar nature, of axis close to the rotation axis (g_1^0 is much larger than the other harmonics); Uranus and Neptune have magnetic fields that are intrinsically much more complex. To provide an idea of the intensity of the magnetic fields, the values of the dipolar moments for the four planets are 4.27 gauss R_J^3, 0.21 gauss R_S^3, 0.23 gauss R_{UI}^3, and 0.133 gauss R_N^3, respectively (Acuna et al., 1983; Connerney et al., 1982; Ness et al., 1986, 1989; see also **Chapter 10.06**).

A true surprise from *Voyager* that has been confirmed by the *Cassini–Huygens* mission is that Saturn's magnetic field is axisymmetric *to the limit of the measurement accuracy*: Saturn's magnetic and rotation axes are perfectly aligned (e.g., Russell and Dougherty, 2010). Voyager measurements indicated nevertheless a clear signature in the radio signal at $10^h39^m22.4^s$ believed to be a consequence of the rotation of the magnetic field. New measurements of a slower spin period of $10^h47^m6^s$ by Cassini (Giampieri et al., 2006; Gurnett et al., 2005) have shown that the kilometric radiation was not directly tied to the period of the magnetic field but resulted from a complex interplay between the spin of the planetary magnetic field and the solar wind (e.g., Cecconi and Zarka, 2005). New periods have been proposed: $10^h32^m35^s$ based on a minimization of the zonal differential rotation (Anderson and Schubert, 2007) and $10^h33^m13^s$ based on the latitudinal distribution of potential vorticity (Read et al., 2009). The problem still stands out.

These magnetic fields must be generated in the conductive parts of the interiors, that is, in metallic hydrogen at radii that

Table 1 Characteristics of the gravity fields and radii

	Jupiter	*Saturn*	*Uranus*	*Neptune*
$M \times 10^{-26}$ (kg)	18.986112(15)[a]	5.68463036(16)[b]	0.8683205(34)[c]	1.0243547861(15)[d]
$R_{eq} \times 10^{-7}$ (m)	7.1492(4)[e]	6.0268(4)[f]	2.5559(4)[g]	2.4766(15)[g]
$R_{pol} \times 10^{-7}$ (m)	6.6854(10)[e]	5.4364(10)[f]	2.4973(20)[g]	2.4342(30)[g]
$\overline{R} \times 10^{-7}$ (m)	6.9894(6)[h]	5.8210(6)[i]	2.5364(10)[i]	2.4625(20)[i]
$\overline{\rho} \times 10^{-3}$ (kg m^{-3})	1.3275(4)	0.6880(2)	1.2704(15)	1.6377(40)
$R_{ref} \times 10^{-7}$ (m)	7.1398[a]	6.0330[b]	2.5559	2.5225[d]
$J_2 \times 10^2$	1.4736(1)[a]	1.629071(27)[b]	0.35160(32)[c]	0.34084(45)[d]
$J_4 \times 10^4$	−5.87(5)[a]	−9.358(28)[b]	−0.354(41)[c]	−0.334(29)[d]
$J_6 \times 10^4$	0.31(20)[a]	0.861(96)[b]		
$P_\omega \times 10^{-4}$ (s)	3.57297(41)[j]	3.83624(47)[j,k]	6.206(4)[l]	5.800(20)[m]
q	0.08923(5)	0.15491(10)	0.02951(5)	0.02609(23)
C/MR_{eq}^2	0.258	0.220	0.230	0.241

The numbers in parentheses are the uncertainty in the last digits of the given value. The value of the gravitational constant used to calculate the masses of Jupiter and Saturn is $G = 6.67259 \times 10^{-11}$ N m^2 kg^{-2} (Cohen and Taylor, 1987). The values of the radii and density correspond to the one bar pressure level (1 bar $= 10^5$ Pa). Gravitational moments are normalized at the reference radius R_{ref}. Only values published in refereed journals are considered.

[a] Campbell and Synnott (1985).

[b] Jacobson et al. (2006).

[c] Anderson et al. (1987).

[d] Jacobson (2009).

[e] Lindal et al. (1981) and Helled and Guillot (2013) derived slightly different values.

[f] Lindal et al. (1985).

[g] Lindal (1992) and Helled et al. (2010) derived slightly different values.

[h] From fourth order figure theory.

[i] $(2R_{eq} + R_{pol})/3$ (Clairaut's approximation).

[j] See also Davies et al. (1986).

[k] This measurement from the *Voyager* era is now in question and values down to 37 955 s have been proposed (see **Section 10.16.2.3**).

[l] Warwick et al. (1986). See however Helled et al. (2010).

[m] Warwick et al. (1989). See however Helled et al. (2010).

are about 80% and 60% of the planetary radius for Jupiter and Saturn, respectively (see **Sections 10.16.3.2** and **10.16.4.1**; e.g., Stanley and Glatzmaier, 2010). Recent models that consistently join the slowly convecting metallic interior with the nonconducting outer molecular envelope dominated by zonal flows result in a mainly dipolar magnetic field similar to the observations and further show that Jupiter's stronger magnetic field and Saturn's broader equatorial jet (see next section) can be interpreted as resulting from the deeper location of the transition region in Saturn (Heimpel and Gómez Pérez, 2011). These explanations remain largely qualitative rather than quantitative and are further complicated by a necessary overforcing of the simulations (see Showman et al., 2011, and next section). The question of why Saturn's magnetic field is much more axisymmetric than Jupiter's remains. Dipolar fields are obtained relatively naturally through the forcing of zonal jets extending down to the conducting region (Guervilly et al., 2012) but why Jupiter differs is unexplained. A possibility is that both fields are nonaxisymmetric at deep levels but that Saturn's is filtered by a more extended helium sedimentation region (Stevenson, 1983), but in practice, a realistic solution yielding Saturn's measured field has not been found (Stanley and Glatzmaier, 2010, and references therein).

Within Uranus and Neptune, the magnetic field is believed to be generated within a layer in which water is in an ionic phase, below about 80% of their total radius (see Redmer et al., 2011, and **Section 10.16.4.2** hereafter). Their complex, multipolar magnetic fields has been thought to be a consequence of a strong stratification and of a dynamo generated in a thin shell (Stanley and Bloxham, 2004). However, this point of view is now challenged by new simulations that generate both planets' magnetic fields and zonal wind structures through a thick shell dynamo (Soderlund et al., 2013). Further work however must involve realistic variations of the interior density and conductivity.

10.16.2.4 Atmospheric Dynamics: Winds and Weather

The atmospheres of all giant planets are evidently complex and turbulent in nature. This can, for example, be seen from the mean zonal winds (inferred from cloud tracking), which are very rapidly varying functions of the latitude (see, e.g., Ingersoll et al., 1995): while some of the regions rotate at the same speed as the interior magnetic field (in the so-called 'system III' reference frame), most of the atmospheres do not. Jupiter and Saturn both have superrotating equators (+100 and +400 m s^{-1} in system III, for Jupiter and Saturn, respectively), and Uranus and Neptune have subrotating equators and superrotating high-latitude jets. Neptune, which receives the smallest amount of energy from the Sun, has the largest peak-to-peak latitudinal variations in wind velocity: about 600 m s^{-1}. It can be noted that, contrary to the case of the strongly irradiated planets to be discussed later, the winds of Jupiter, Saturn, Uranus, and Neptune are significantly slower than the planet itself under its own spin (from 12.2 km s^{-1} for Jupiter to 2.6 km s^{-1} for Neptune, at the equator).

It is not yet clear whether the observed winds are driven from the bottom or from the top. The first possibility is that surface winds are related to motions in the planets' interiors, which, according to the Taylor–Proudman theorem, should be confined by the rapid rotation to the plane perpendicular to the axis of rotation (Busse, 1978). This is now backed by simulations in the anelastic limit (i.e., accounting for compressibility), which show that the outcome strongly depends on the density stratification in the interior. A small density contrast (as expected in Jupiter and Saturn) leads to equatorial superrotation, whereas for a large one (as expected for Uranus and Neptune), the equatorial jet tends to subrotate (Gastine et al., 2013; Glatzmaier et al., 2009). However, the application of these numerical results to the true conditions prevailing in the giant planets requires an extrapolation over at least six orders of magnitude (Showman et al., 2011). The second possibility (not exclusive) is that winds are driven from the top by the injection of turbulence at the cloud level, which can also lead to the correct winds for the four giant planets (Lian and Showman, 2010; Liu and Schneider, 2011).

Information on the gravity field of the planets can be used to constrain the interior rotation profile, as in the case of Uranus and Neptune whose observed jets appear to only extend to the outer 0.15% and 0.20% of the mass (corresponding to pressures of 2 and 4 kbar) for Uranus and Neptune, respectively (Kaspi et al., 2013). The method is promising with the perspective of the Juno measurements at Jupiter (Liu et al., 2013).

Our giant planets also exhibit planetary-scale to small-scale storms with very different temporal variations. For example, Jupiter's Great Red Spot is a 12 000 km diameter anticyclone found to have lasted for at least 300 years (e.g., Simon-Miller et al., 2002). Storms developing over the entire planet have even been observed on Saturn (Sanchez-Lavega et al., 1996). Uranus and Neptune's storm system has been shown to have been significantly altered since the Voyager era (de Pater et al., 2011; Hammel et al., 2005; Rages et al., 2002). On Jupiter, small-scale storms related to cumulus-type cloud systems have been observed (e.g., Gierasch et al., 2000; Hueso et al., 2002), and lightning strikes have been monitored by Galileo (e.g., Little et al., 1999). These represent only a small arbitrary subset of the work concerning the complex atmospheres of these planets.

It is tempting to extrapolate these observations to the objects outside our solar system as well. However, two features governing the weather in these are not necessarily present for exoplanets (e.g., Guillot, 1999b): their rapid rotation and the presence of abundant condensing species and in particular one, water, whose latent heat can fuel powerful storms. But as we will see briefly in **Section 10.16.2.10** theoretical models for exoplanets are now complemented by measurements of wind speeds and of global temperature contrasts, offering the perspective of a global approach to planetary weather and atmospheric dynamics.

10.16.2.5 Energy Balance and Atmospheric Temperature Profiles

Jupiter, Saturn, and Neptune are observed to emit more energy than they receive from the Sun (see **Table 2**). The case of Uranus is less clear. Its intrinsic heat flux $F_{\rm int}$ is significantly smaller than that of the other giant planets. With this caveat, all four giant planets can be said to emit more energy than they receive from the Sun. Hubbard (1968) showed in the case of Jupiter that this can be explained simply by the progressive contraction and cooling of the planets.

A crucial consequence of the presence of an intrinsic heat flux is that it requires high internal temperatures ($\sim$10 000 K or more) and that consequently the giant planets are *fluid* (not solid) (Hubbard, 1968; see also Hubbard et al., 1995). Another consequence is that they are essentially convective and that their interior temperature profile is close to *adiabats*. We will come back to this in more detail.

The deep atmospheres (more accurately tropospheres) of the four giant planets are indeed observed to be close to adiabats, a result first obtained by spectroscopic models (Trafton, 1967), then verified by radio-occultation experiments by the Voyager spacecrafts, and by the in situ measurement from the Galileo probe (**Figure 3**). The temperature profiles show a temperature minimum, in a region near 0.2 bar called the tropopause. At higher altitudes, in the stratosphere, the temperature gradient is negative (increasing with decreasing pressure). In the regions that we will be mostly concerned with, in the troposphere and in the deeper interior, the temperature always increases with depth. It can be noticed that the slope of the temperature profile in **Figure 3** becomes almost constant

Table 2 Energy balance as determined from Voyager IRIS data

	Jupiter	*Saturn*	*Uranus*	*Neptune*
Absorbed power (10^{16} J s^{-1})	50.14 ± 2.48	11.14 ± 0.50	0.526 ± 0.037	0.204 ± 0.019
Emitted power (10^{16} J s^{-1})	83.65 ± 0.84	19.77 ± 0.32	0.560 ± 0.011	0.534 ± 0.029
Intrinsic power (10^{16} J s^{-1})	33.5 ± 2.6	8.63 ± 0.60	$0.034^{+0.038}_{-0.034}$	0.330 ± 0.035
Intrinsic flux (J s^{-1} M^{-2})	5.44 ± 0.43	2.01 ± 0.14	$0.042^{+0.047}_{-0.042}$	0.433 ± 0.046
Bond albedo	0.343 ± 0.032	0.342 ± 0.030	0.300 ± 0.049	0.290 ± 0.067
Effective temperature (K)	124.4 ± 0.3	95.0 ± 0.4	59.1 ± 0.3	59.3 ± 0.8
1 Bar temperature[a] (K)	165 ± 5	135 ± 5	76 ± 2	72 ± 2

[a]Lindal (1992).

Source: Pearl JC and Conrath BJ (1991)

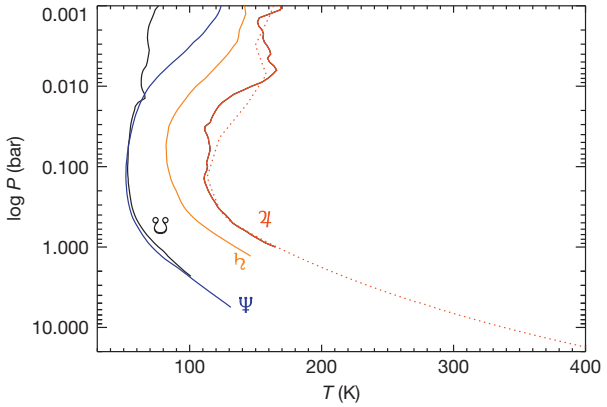


Figure 3 Atmospheric temperatures as a function of pressure for Jupiter, Saturn, Uranus, and Neptune, as obtained from Voyager radio-occultation experiments (see Lindal, 1992). The dotted line corresponds to the temperature profile retrieved by the Galileo probe, down to 22 bar and a temperature of 428 K (Seiff et al. 1998).

when the atmosphere becomes convective, at pressures of a fraction of a bar, in the four giant planets.

It should be noted that the 1 bar temperatures listed in Table 2 and the profiles shown in Figure 3 are retrieved from radio-occultation measurements using a helium–hydrogen ratio, which, at least in the case of Jupiter and Saturn, was shown to be incorrect. The new values of Y are found to lead to increased temperatures by $\sim$5 K in Jupiter and $\sim$10 K in Saturn (see Guillot, 1999a). However, the Galileo probe found a 1 bar temperature of 166 K (Seiff et al., 1998) and generally a good agreement with the Voyager radio-occultation profile with the wrong He/H_2 value.

When studied at low spatial resolution, it is found that all four giant planets, in spite of their inhomogeneous appearances, have a rather uniform brightness temperature, with pole-to-equator latitudinal variations limited to a few kelvins (e.g., Ingersoll et al., 1995). However, in the case of Jupiter, some small regions are known to be very different from the average of the planet. This is the case of hot spots, which cover about 1% of the surface of the planet at any given time but contribute to most of the emitted flux at 5 μm, due to their dryness (absence of water vapor) and their temperature brightness which can, at this wavelength, peak to 260 K.

10.16.2.6 Atmospheric Compositions

In fluid planets, the distinction between the atmosphere and the interior is not obvious. We name 'atmosphere' the part of the planet that can directly exchange radiation with the exterior environment. This is also the part that is accessible by remote sensing. It is important to note that the continuity between the atmosphere and the interior does not guarantee that compositions measured in the atmosphere can be extrapolated to the deep interior, even in a fully convective environment: Processes such as phase separations (e.g., Fortney and Hubbard, 2003; Salpeter, 1973; Stevenson and Salpeter, 1977b), phase transitions (e.g., Hubbard, 1989), chemical reactions (e.g., Fegley and Lodders, 1994), and cloud formation (e.g., Rossow, 1978) can occur and decouple the surface and interior compositions. Furthermore, imperfect mixing may also occur, depending on the initial conditions (e.g., Stevenson, 1985).

The conventional wisdom is however that these processes are limited to certain species (e.g., helium) or that they have a relatively small impact on the global abundances, so that the hydrogen–helium envelopes may be considered relatively uniform, from the perspective of the global abundance in heavy elements. An important caveat is that measurements must probe deeper than the condensation altitude for any volatile (e.g., ammonia and water). We first discuss measurements made in the atmosphere before inferring interior compositions from interior and evolution models.

10.16.2.6.1 Hydrogen and helium

The most important components of the atmospheres of our giant planets are also among the most difficult to detect: H_2 and He have a zero dipolar moment and hence absorb very inefficiently visible and infrared light. Absorption in the infrared becomes important only at high pressures as a result of collision-induced absorption (e.g., Borysow et al., 1997). On the other hand, lines due to electronic transitions correspond to very high altitudes in the atmosphere and bear little information on the structure of the deeper levels. The only robust result concerning the abundance of helium in a giant planet is by in situ measurement by the Galileo probe in the atmosphere of Jupiter (von Zahn et al., 1998). The helium mole fraction (i.e., number of helium atoms over the total number of species in a given volume) is $q_{He} = 0.1359 \pm 0.0027$. The helium mass mixing ratio Y (i.e., mass of helium atoms over total mass) is constrained by its ratio over hydrogen, $X{:}Y/(X+Y) = 0.238 \pm 0.05$. This ratio is by coincidence that found in the Sun's atmosphere, but because of helium sedimentation in the Sun's radiative zone, it was larger in the protosolar nebula: $Y_{proto} = 0.275 \pm 0.01$ and $(X+Y)_{proto} \approx 0.98$ (e.g., Bahcall et al., 1995). Less helium is therefore found in the atmosphere of Jupiter than inferred to be present when the planet formed. We will discuss the consequences of this measurement later: let us mention that the explanation invokes helium settling due to a phase separation in the interiors of massive and cold giant planets.

Helium is also found to be depleted compared to the protosolar value in Saturn's atmosphere. However, in this case, the analysis is complicated by the fact that Voyager radio-occultations combined with the far-IR sounding (to separate effects of helium from that of temperature) led to a wrong value for Jupiter when compared to the Galileo probe data and hence are suspect for the other planets. The current adopted value from IR data only is now $Y = 0.18 - 0.25$ (Conrath and Gautier, 2000), in agreement with values predicted by interior and evolution models (Guillot, 1999b; Hubbard et al., 1999).

Finally, as shown in Table 3 hereafter, Uranus and Neptune are found to have near-protosolar helium mixing ratios, but with considerable uncertainty.

10.16.2.6.2 Heavy elements

The abundance of elements other than hydrogen and helium (that we will call hereafter 'heavy elements') bears crucial information for the understanding of the processes that led to the formation of these planets. Table 3 summarizes the present

Table 3 Elemental abundances measured in the tropospheres of the giant planets

Element	*Carrier*	*Abundance ratio/H*[a]	*Protosun*[b]	*Planet/protosun*	*Method*
Jupiter					
He/H	He	$(7.85 \pm 0.18) \times 10^{-2}$	9.69×10^{-2}	0.810 ± 0.019	Galileo/GPMS[c]
C/H	CH_4	$(1.185 \pm 0.019) \times 10^{-3}$	2.75×10^{-4}	4.31 ± 0.07	Galileo/GPMS[d]
N/H	NH_3	$(3.3 \pm 1.3) \times 10^{-4}$	8.19×10^{-5}	4.05 ± 1.55	Galileo/GPMS[d]
O/H	H_2O[e]	$(1.49^{+0.98}_{-0.68}) \times 10^{-4}$	6.06×10^{-4}	$0.25^{+0.16}_{-0.11}$	Galileo/GPMS at 19 bar[d]
S/H	H_2S	$(4.5 \pm 1.1) \times 10^{-5}$	1.55×10^{-5}	2.88 ± 0.68	Galileo/GPMS[d]
Ne/H	Ne	$(1.20 \pm 0.12) \times 10^{-5}$	1.18×10^{-4}	0.10 ± 0.01	Galileo/GPMS[f]
Ar/H	Ar	$(9.10 \pm 1.80) \times 10^{-6}$	3.58×10^{-6}	2.54 ± 0.50	Galileo/GPMS[f]
Kr/H	Kr	$(4.65 \pm 0.85) \times 10^{-9}$	2.15×10^{-9}	2.16 ± 0.40	Galileo/GPMS[f]
Xe/H	Xe	$(4.45 \pm 0.85) \times 10^{-10}$	2.11×10^{-10}	2.11 ± 0.40	Galileo/GPMS[f]
P/H	PH_3[e]	$(1.11 \pm 0.06) \times 10^{-6}$	3.20×10^{-7}	3.45 ± 0.18	Cassini/CIRS[g]
Ge/H	GeH_4[e]	$(4.1 \pm 1.2) \times 10^{-10}$	4.44×10^{-9}	0.09 ± 0.03	Voyager/IRIS[h]
As/H	AsH_3[e]	$(1.3 \pm 0.6) \times 10^{-10}$	2.36×10^{-10}	0.54 ± 0.27	Ground/IR[i]
Saturn					
He/H	He	$(6.75 \pm 1.25) \times 10^{-2}$	9.69×10^{-2}	0.70 ± 0.13	Voyager/IRIS[j]
C/H	CH_4	$(2.67 \pm 0.11) \times 10^{-3}$	2.75×10^{-4}	9.72 ± 0.41	Cassini/CIRS[k]
N/H	NH_3[e]	$(2.27 \pm 0.57) \times 10^{-4}$	8.19×10^{-5}	2.77 ± 0.69	Cassini/VIMS[l]
S/H	H_2S	$(1.25 \pm 0.17) \times 10^{-4}$	1.55×10^{-5}	8.08 ± 1.10	Ground/radio[m]
P/H	PH_3[e]	$(4.65 \pm 0.32) \times 10^{-6}$	3.20×10^{-7}	14.5 ± 1.0	Cassini/CIRS[g]
		$(1.76 \pm 0.17) \times 10^{-6}$	3.20×10^{-7}	5.49 ± 0.53	Cassini/VIMS[k]
		$(4.0^{+1.7}_{-1.1}) \times 10^{-6}$	3.20×10^{-7}	$12.4^{+5.3}_{-3.5}$	Ground/IR[n]
Ge/H	GeH_4[e]	$(2.3 \pm 2.3) \times 10^{-10}$	4.44×10^{-9}	0.05 ± 0.05	Ground/IR[n]
As/H	AsH_3[e]	$(1.25 \pm 0.17) \times 10^{-9}$	2.36×10^{-10}	5.33 ± 0.73	Cassini/VIMS[k]
		$(1.71 \pm 0.57) \times 10^{-9}$	2.36×10^{-10}	7.3 ± 2.4	Ground/IR[n]
Uranus					
He/H	He	$(9.0 \pm 2.0) \times 10^{-2}$	9.69×10^{-2}	0.93 ± 0.20	Voyager/IRIS + OCCULT[o]
C/H	CH_4[e]	$(2.36 \pm 0.30) \times 10^{-2}$	2.75×10^{-4}	85.9 ± 10.7	Hubble/STIS[p]
S/H	H_2S[e]	$(3.2 \pm 1.6) \times 10^{-4}$	1.55×10^{-5}	21.0 ± 10.5	Ground/RADIO[q]
Neptune					
He/H	He	$(1.17 \pm 0.20) \times 10^{-1}$	9.69×10^{-2}	1.21 ± 0.20	Voyager/IRIS + OCCULT[r]
C/H	CH_4[e]	$(1.85 \pm 0.43) \times 10^{-2}$	2.75×10^{-4}	67.5 ± 15.8	Ground/IR[s]
		$(2.47 \pm 0.62) \times 10^{-2}$	2.75×10^{-4}	89.9 ± 22.5	Hubble/STIS[t]
S/H	H_2S[e]	$(3.2 \pm 1.6) \times 10^{-4}$	1.55×10^{-5}	21.0 ± 10.5	Ground/RADIO[q]

[a]Abundance ratios r are measured with respect to atomic hydrogen. In these atmospheres dominated by molecular hydrogen and helium, mole fractions f are found by $f = 2r/(1 + r_{He})$ where r_{He} is the He/H abundance ratio.

[b]Protosolar abundances from Lodders et al. (2009).

[c]von Zahn et al. (1998).

[d]Wong et al. (2004).

[e]Species that condense or are in chemical disequilibrium, that is, with vertical/horizontal variations of their concentration. The global elemental abundances are estimated from the maximum measured mixing ratio, but like in the case of H_2O in Jupiter (believed to correspond to the measurement in a dry downdraft), they may only be lower limits to the bulk abundance.

[f]Atreya et al. (2003).

[g]Fletcher et al. (2009a).

[h]Kunde et al. (1982).

[i]Noll et al. (1990).

[j]Conrath and Gautier (2000).

[k]Fletcher et al. (2009b).

[l]Fletcher et al. (2011).

[m]Briggs and Sackett (1989).

[n]Noll and Larson (1991).

[o]Conrath et al. (1987).

[p]Sromovsky et al. (2011).

[q]de Pater et al. (1991).

[r]Conrath et al. (1991).

[s]Baines and Smith (1990).

[t]Karkoschka and Tomasko (2011).

situation after in situ measurements in Jupiter by the Galileo probe, as well as spectroscopic measurements from spacecraft and from the ground for the other planets.

The elemental abundances in the giant planets' atmospheres are most usefully compared to those in the Sun since they all originated from the protosolar disk. The solar abundances have seen very significant revisions in the past decade because it has been realized that convective motions in the Sun's atmosphere affect spectral lines more extensively than what was previously thought. It is not yet clear at this date whether the solar abundances have converged. Furthermore, as discussed for helium, heavy elements gradually settle toward the Sun's interior so that a proper reference for the giant planets is not the solar atmosphere today, but its value 4.5 billion years ago that is model-dependent. **Table 3** provides the values obtained for the protosun by Lodders et al. (2009). These are used as reference without accounting for their uncertainties.

The most abundant heavy elements in the envelopes of our four giant planets are O (presumably) and C, N, and S. It is possible to model the chemistry of gases in the tropospheres from the top of the convective zone down to the 2000 K temperature level (Fegley and Lodders, 1994). Models conclude that, whatever the initial composition in these elements of planetesimals that collapsed with hydrogen onto Jupiter and Saturn cores during the last phase of the planetary formation, C in the upper tropospheres of the giant planets is mainly in the form of gaseous CH_4, N in the form of NH_3, S in the form of H_2S, and O in the form of H_2O. All these gases but methane in Jupiter and Saturn condense in the upper troposphere and vaporize at deeper levels when the temperature increases. Noble gases do not condense even at the tropopauses of Uranus and Neptune, the coldest regions in these atmospheres.

Jupiter is the planet that has been best characterized, thanks to the measurements of the Galileo atmospheric probe that precisely measured the abundances of He, Ne, Ar, Kr, Xe, CH_4, NH_3, H_2S, and H_2O down to pressures around 22 bars. As helium, neon was found to be depleted compared to the protosolar value, in line with theoretical predictions that this atom would fall in with the helium droplets (Roulston and Stevenson, 1995; Wilson and Militzer, 2010; **Section 10.16.3.2.2**). C, N, and S were found to be supersolar by a factor 2.5–4.5 (Wong et al., 2004), which was not unexpected because condensation of nebula gases results in enriching icy grains and planetesimals. The surprise came from Ar, Kr, and Xe, which were expected to be solar because they are difficult to condense but turned out to be supersolar by a factor $\sim$2 (Owen et al., 1999; Wong et al., 2004).

H_2O is difficult to measure in all four giant planets because of its condensation being relatively deep. It was hoped that the Galileo probe would provide a measurement of its deep abundance, but the probe fell into one of Jupiter's 5 μm hot spots, now believed to be a dry region mostly governed by downwelling motions (e.g., Showman and Ingersoll, 1998). As a result and although the probe provided measurements down to 22 bars, well below water's canonical 5 bar cloud base, it is believed that this measurement of a water abundance equal to a fraction of the solar value is only a lower limit. An indirect determination comes from the measurement of the disequilibrium species CO that has to be transported fast from the deep levels where H_2O and CH_4 tend to form more CO (and H_2). This predicts a mostly solar to slightly supersolar (by a factor 2) abundance of O in Jupiter (Visscher and Moses, 2011) and much larger enrichments in Neptune (Lodders and Fegley, 1994). This however depends crucially on the reaction network and somewhat on assumptions on mixing, both of which are not well known. The abundance of oxygen, the most abundant element in the universe after hydrogen and helium and a crucial planetary building block, is essentially unknown for what concerns our four giant planets.

Three other species can help us probe the bulk elemental abundance inside Jupiter, although with larger difficulties perhaps because they are not necessarily in chemical equilibrium at the levels where they are detected and their measured abundances are thus not necessarily representative of their bulk abundance: these are PH_3, GeH_4, and AsH_3. All three were detected remotely rather than in situ. The first one is clearly supersolar in Jupiter, with an enrichment in between that measured for C and S. GeH_4 is clearly subsolar, but this is not surprising because of condensation into solid Ge and GeS (Fegley and Lodders, 1994). The same chemical models would predict that the measured abundance of AsH_3 should be close to its bulk abundance. The measured abundance therefore could be interpreted as a subsolar bulk abundance of As but with considerable uncertainty.

Table 3 shows that Saturn's atmosphere is more enriched in heavy elements than Jupiter. Unfortunately, unlike Jupiter, no in situ measurement has been performed in this planet and we can only rely on remote sensing. But we can confidently assess that the abundances of C, S, P, and As are significantly higher than in Jupiter. Saturn has about twice more C (as CH_4) and S (as H_2S) than Jupiter, for a given mass of atmosphere. The situation for PH_3 is unclear, because it is highly variable both vertically and latitudinally: the enrichment could be only slightly more than in Jupiter to more than four times that value (Fletcher et al., 2009a, 2011). Note that in the presence of horizontal variability (as for this molecule), **Table 3** indicates the maximum abundances measured – which should be closer to the bulk abundance, except if there exist mechanisms to preferentially trap certain species. The enrichment in N (as NH_3) is smaller than in Jupiter. This may be due to its condensation deeper as NH_4SH (Gulkis et al., 1978), although it does not explain why Jupiter and Saturn would be that different in that respect. The enrichment in As (as AsH_4) is considerably larger than in Jupiter, which is also a mystery (Fletcher et al., 2009b). At least, GeH_4 appears to be of equally low abundance in both planets, but this is probably more related to its condensation than to its bulk abundance.

Finally, Uranus and Neptune provide all signs of a significant enrichment in heavy elements, even though very few elements have been detected. Methane is the most important one, although the fact that it condenses in these planets complicates the interpretation of the spectroscopic measurements. Large-scale variations with latitude are observed, in particular in Uranus (Sromovsky et al., 2011), but less so in Neptune (Karkoschka and Tomasko, 2011). However, the deep abundances provided in **Table 3** are very similar for both planets, with a $\sim$90 times solar enrichment. This is much higher than

the ~30 times solar enrichments discussed in past reviews (e.g., Gautier et al., 1995) for two reasons: one is a decrease of the protosolar abundance itself. The other is the fact that it relied on spectroscopic measurements probing higher atmospheric levels affected by methane condensation (see Karkoschka and Tomasko, 2011).

The other key species detected in Uranus and Neptune thanks to ground-based radio observations is H_2S, which points to a 10–30 enrichment in sulfur (de Pater et al., 1991). Because this element condenses at even greater depths than methane, the bulk abundance of S in these planets may be larger if the global circulation is indeed important down to the deep levels probed by the radio waves. The measurement is however a difficult one, with other potential absorbers affecting the smooth microwave spectra yielding degenerate solutions.

Overall, the global picture that can be drawn is that of an increase of the abundance of heavy elements compared to the solar value with increasing distance to the Sun, from Jupiter that shows a ~2–4 enrichment, Saturn that shows a 3–10 one, and Uranus and Neptune that are enriched by a factor ~90 in carbon and by at least 10–30 in sulfur. In spite of their different atmospheric dynamics, and with the present accuracy of the measurements, the two ice giants have very similar abundances.

10.16.2.7 Isotopic Ratios

The measurement of isotopic ratios in planetary atmospheres is a powerful tool to understand their origin. Table 4 provides the ensemble of isotopic ratios measured in our giant planets and a comparison to their values in the Sun.

Of course, the Galileo probe and its onboard mass spectrometer have provided us a strikingly clear picture of Jupiter's atmosphere: it is directly formed from the same material as our Sun, with isotopic ratios that are, to the accuracy of the measurements, indistinguishable (i.e., within 2 sigma) from the solar values and for elements as diverse as D, He, C, N, Ne, Ar, and Xe with as many as six isotopes measured. This was expected because indeed Jupiter's composition is globally similar to that of the Sun, but given the fact that the abundances of elements are far from being Sun-like, it is perhaps surprising to find such a good match! By extension, this applies to Saturn although only the deuterium to hydrogen and $^{13}C/^{12}C$ isotopic ratios could be measured by remote spectroscopic observations. This confirms that the atmospheres and envelopes of Jupiter and Saturn originated from the same material that formed the Sun and that mechanisms leading to isotopic fractionation (e.g., atmospheric evaporation) were of limited importance.

In the case of Uranus and Neptune, only the deuterium to hydrogen ratio was measured, from the ground in the infrared (Irwin et al., 2014) and most precisely by recent far-infrared spectroscopy from Herschel (Feuchtgruber et al., 2013). Interestingly, it is about twice larger than the protosolar value and a factor two to six times smaller than the D/H value in comets. Given our present knowledge of the interiors of Uranus and Neptune, Feuchtgruber et al. (2013) concluded that either these planets contain much more rocks than expected or the ices in their interior have not been fully mixed.

Table 4 Isotopic ratios measured in the tropospheres of the giant planets

Isotope	*Isotopic ratio*	*Protosun*[a]	*Planet/protosun*	*Comments*
Jupiter				
D/H	$(2.25\pm0.35)\times10^{-5}$	1.94×10^{-5}	1.16 ± 0.18	ISO/SWS[b]
$^{3}He/^{4}He$	$(1.66\pm0.06)\times10^{-4}$	1.66×10^{-4}	1.00 ± 0.03	Galileo/GPMS[c]
$^{13}C/^{12}C$	$(1.08\pm0.05)\times10^{-2}$	1.12×10^{-2}	1.04 ± 0.05	Galileo/GPMS[c]
$^{15}N/^{14}N$	$(2.30\pm0.30)\times10^{-3}$	2.27×10^{-3}	0.99 ± 0.13	Galileo/GPMS[c]
$^{22}Ne/^{20}Ne$	$(7.7\pm1.2)\times10^{-2}$	7.35×10^{-2}	0.96 ± 0.15	Galileo/GPMS[c]
$^{38}Ar/^{36}AR$	$(1.79\pm0.08)\times10^{-1}$	1.82×10^{-1}	1.02 ± 0.05	Galileo/GPMS[c]
$^{128}Xe/Xe$	$(1.80\pm0.20)\times10^{-2}$	2.23×10^{-2}	1.24 ± 0.14	Galileo/GPMS[c]
$^{129}Xe/Xe$	$(2.85\pm0.21)\times10^{-1}$	2.75×10^{-1}	0.96 ± 0.07	Galileo/GPMS[c]
$^{130}Xe/Xe$	$(3.80\pm0.50)\times10^{-2}$	4.38×10^{-2}	1.15 ± 0.15	Galileo/GPMS[c]
$^{131}Xe/Xe$	$(2.03\pm0.18)\times10^{-1}$	2.18×10^{-1}	1.07 ± 0.10	Galileo/GPMS[c]
$^{132}Xe/Xe$	$(2.90\pm0.20)\times10^{-1}$	2.64×10^{-1}	0.91 ± 0.06	Galileo/GPMS[c]
$^{134}Xe/Xe$	$(9.10\pm0.70)\times10^{-2}$	9.66×10^{-2}	1.06 ± 0.08	Galileo/GPMS[c]
Saturn				
D/H	$(1.60\pm0.20)\times10^{-5}$	1.94×10^{-5}	0.83 ± 0.11	Cassini/CIRS[d]
	$(1.70^{+0.75}_{-0.45})\times10^{-5}$	1.94×10^{-5}	$0.88^{+0.39}_{-0.23}$	ISO/SWS[b]
$^{13}C/^{12}C$	$(1.09\pm0.10)\times10^{-2}$	1.12×10^{-2}	1.03 ± 0.09	Cassini/CIRS[d]
Uranus				
D/H	$(4.40\pm0.40)\times10^{-5}$	1.94×10^{-5}	2.27 ± 0.21	Herschel/PACS[e]
Neptune				
D/H	$(4.10\pm0.40)\times10^{-5}$	1.94×10^{-5}	2.11 ± 0.21	Herschel/PACS[e]

[a]Protosolar abundances from Lodders et al. (2009), except $^{15}N/^{14}N$, which is corrected by Marty et al. (2011).
[b]Lellouch et al. (2001).
[c]Atreya et al. (2003).
[d]Fletcher et al. (2009b).
[e]Feuchtgruber et al. (2013).

10.16.2.8 Moons and Rings

A discussion of our giant planets motivated by the opportunity to extrapolate the results to objects outside our solar system would be incomplete without mentioning the moons and rings that these planets all possess (see Chapters 10.08, 10.17, and 10.18). First, the satellites/moons can be distinguished from their orbital characteristics as regular or irregular. The first ones have generally circular, prograde orbits. The latter tend to have eccentric, extended, and/or retrograde orbits.

These satellites are numerous: After the Voyager era, Jupiter was known to possess 16 satellites, Saturn 18, Uranus 20, and Neptune 8. Recent extensive observation programs have seen the number of satellites increase considerably, with a growing list of satellites presently reaching 62, 56, 27, and 13 for Jupiter, Saturn, Uranus, and Neptune, respectively. All of the new satellites discovered since Voyager are classified as irregular.

The presence of regular and irregular satellites is due in part to the history of planet formation. It is believed that the regular satellites have mostly been formed in the protoplanetary subnebulae that surrounded the giant planets (at least Jupiter and Saturn) at the time when they accreted their envelopes. On the other hand, the irregular satellites are thought to have been captured by the planet. This is, for example, believed to be the case of Neptune's largest moon, Triton, which has a retrograde orbit.

A few satellites stand out by having relatively large masses: it is the case of Jupiter's Io, Europa, Ganymede, and Callisto; of Saturn's Titan; and of Neptune's Triton. Ganymede is the most massive of them, being about twice the mass of our Moon. However, compared to the mass of the central planet, these moons and satellites have very small weights: 10^{-4} and less for Jupiter, 1/4000 for Saturn, 1/25 000 for Uranus, and 1/4500 for Neptune. All these satellites orbit relatively closely to their planets. The farthest one, Callisto revolves around Jupiter in about 16 Earth days.

The four giant planets also have rings, whose material is probably constantly resupplied from their satellites. The ring of Saturn stands out as the only one directly visible with binoculars. In this particular case, its enormous area allows it to reflect a sizable fraction of the stellar flux arriving at Saturn and makes this particular ring as bright as the planet itself. The occurrence of such rings would make the detection of extrasolar planets slightly easier, but it is yet unclear how frequent they can be and how close to the stars rings can survive both the increased radiation and tidal forces.

10.16.2.9 Seismology

The best way to directly probe planetary (or stellar) interiors is through seismology, that is, by measuring the spectrum of waves propagating through the interior. Our knowledge of the interior structure of the Earth, the Sun, and even other stars is largely due to the ability to detect the oscillations of these objects. Because Jupiter (and by extent the other giant planets in our solar system) is similar in composition and density to the Sun, the possibility to detect pressure waves in their atmosphere has been proposed in the 1970s. In particular, Vorontsov et al. (1976) showed that waves with periods smaller than about 10 min would be trapped and reflected downward in Jupiter's and Saturn's atmospheres, creating the possibility for resonant waves similar to those observed in the Sun to exist. On the theoretical side, the possibility of how these waves may be excited has however remained problematic (e.g., Bercovici and Schubert, 1987).

After two decades of promising but slow progress, the case for the existence of detectable free oscillations of the giant planets has recently taken a new turn. First, using ground-based Doppler imaging of Jupiter, Gaulme et al. (2011) detected an oscillation pattern with frequencies between 0.8 and 2 mHz and a characteristic spacing of the peak of 155.3 ± 2.2 μHz, in agreement with theoretical models. Separately, Hedman and Nicholson (2013) confirmed that waves observed by the Cassini spacecraft in Saturn's rings cannot be caused by satellites and therefore must result from oscillations in Saturn, as had been proposed by Marley and Porco (1993).

Further observations are required to better characterize the oscillations of these planets and start using them as probes of the planetary interiors (e.g., Jackiewicz et al., 2012). This is very promising however and should lead to a revolution in our understanding of the giant planets.

10.16.2.10 Exoplanets

Huge progress has been made in the field of extrasolar planets since the detection of the first giant planet orbiting a solar-type star by Mayor and Queloz (1995). As shown in Figure 4, more than a thousand planets are known at the time of this chapter, and importantly, more than four hundred planets that transit their star at each orbital revolution have been identified (Schneider et al., 2011; Wright et al., 2011). These transiting planets are especially interesting because of the possibility to

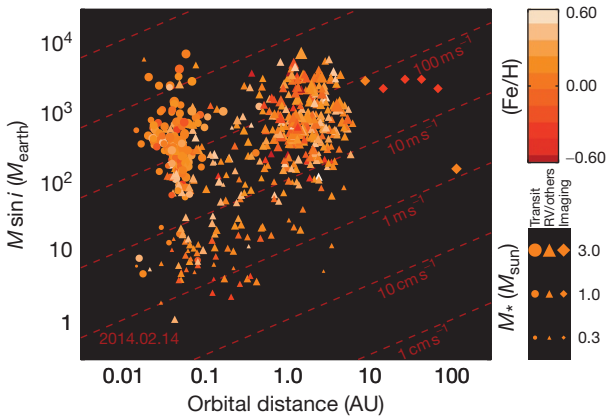


Figure 4 Masses and orbital distances of the extrasolar planets discovered by early 2014. Planets detected in transit are shown with circles, those detected by direct imaging with diamonds. All other systems (mostly detected by radial velocimetry) are shown as triangles. In the case of planets discovered by radial velocimetry, the displayed masses correspond to masses multiplied by the sine of the unknown inclination. The size of the symbols is proportional to the mass of the parent star (from 0.1 to 3.1 stellar masses). The color (from reddish to white) is proportional to the stellar metallicity (from [Fe/H] = −0.8 to 0.6). The radial velocimetry thresholds from 1 cm s^{-1} to 10 km s^{-1} are indicated as dashed lines.

measure both their mass and size and thus obtain constraints on their global composition.

In spite of their particular location just a few stellar radii away from their stars, the transiting giant planets that have been discovered bear some resemblance with their solar system cousins in the sense that they are also mostly made of hydrogen and helium (e.g., Baraffe et al., 2005; Burrows et al., 2000; Guillot, 2005). They are, however, much hotter due to the intense irradiation that they receive.

Although obtaining direct information on these planets represents a great observational challenge, several key steps have been accomplished: Atomic sodium, predicted to be detectable (Seager and Sasselov, 2000), has indeed been detected by transit spectroscopy(Charbonneau et al., 2002) early on and a tentative abundance measured in planet HD 209458 b. According to Sing et al. (2009), it appears to be oversolar by a factor ~ 2 at pressures deeper than about ~ 3 mbar and undersolar above that level (see also Vidal-Madjar et al., 2011). Hydrodynamically, escaping species (including hydrogen, oxygen, carbon, nitrogen, and heavier ions) have also been detected around the brightest hot Jupiters (e.g., Bourrier et al., 2013; Fossati et al., 2010; Vidal-Madjar et al., 2003). A theoretical study of the atmospheric dynamics of hot Jupiters (Showman and Guillot, 2002) predicted strong day–night temperature variations (up to 100s K), fast km s^{-1} zonal jets, and a displacement of the hottest point west of the substellar point. These have been confirmed by observations of transiting and nontransiting planets in the infrared (Harrington et al., 2006; Knutson et al., 2007) and by Doppler imaging of planetary CO lines (Snellen et al., 2010).

Unfortunately, the list of chemical species thought to have been detected (see Seager and Deming, 2010, for a review) has dwindled in recent years due to the realization that instrumental effects could mimic spectral signatures (e.g., Crouzet et al., 2012; Désert et al., 2009; Gibson et al., 2011), due to new observations with a better instrument (e.g., Deming et al., 2013) and generally because of the unexpected prevalence of hazes in these close-in exoplanets (e.g., Pont et al., 2013; Sing et al., 2009). After examination, claims of a high C/O ratio in some of these atmospheres also appear to be highly uncertain (Crossfield et al., 2012).

In any case, in spite of the hiccups, there is obviously a big potential for growth in this young field, and the comparison between fine observations made for the giant planets in our solar system and the more crude but also more statistically significant data obtained for planets around other stars promise to be extremely fruitful to better understand these objects.

10.16.3 The Calculation of Interior and Evolution Models

10.16.3.1 Basic Equations

The structure and evolution of a giant planet is governed by the following hydrostatic, thermodynamic, mass conservation, and energy conservation equations:

$$\frac{\partial P}{\partial r} = -\rho g \qquad [4]$$

$$\frac{\partial T}{\partial r} = \frac{\partial P}{\partial r}\frac{T}{P}\nabla_T \qquad [5]$$

$$\frac{\partial m}{\partial r} = 4\pi r^2 \rho \qquad [6]$$

$$\frac{\partial L}{\partial r} = 4\pi r^2 \rho\left(\dot{\varepsilon} - T\frac{\partial S}{\partial t}\right) \qquad [7]$$

where P is the pressure, ρ the density, and $g = Gm/r^2$ the gravity (m is the mass, r the radius, and G the gravitational constant). The temperature gradient $\nabla_T \equiv (d\ln T/d\ln P)$ depends on the process by which the internal heat is transported. L is the intrinsic luminosity, t is the time, S is the specific entropy (per unit mass), and $\dot{\varepsilon}$ accounts for the sources of energy due, for example, to radioactivity or more importantly nuclear reactions. Generally, it is a good approximation to assume $\dot{\varepsilon} \sim 0$ for objects less massive than ~ 13 M_J, that is, too cold to even burn deuterium (but we will see that in certain conditions, this term may be useful, even for low-mass planets).

The boundary condition at the center is trivial: $r = 0$; ($m = 0$, $L = 0$). The external boundary condition is more difficult to obtain because it depends on how energy is transported in the atmosphere. One possibility is to use the Eddington approximation and to write (e.g., Chandrasekhar, 1939): $r = R$; ($T_0 = T_{\text{eff}}$, $P_0 = 2/3\, g/\kappa$), where T_{eff} is the effective temperature (defined by $L = 4\pi R\sigma T_{\text{eff}}^4$, with σ being the Stefan–Boltzmann constant) and κ is a mean opacity. Note, for example, that in the case of Jupiter, $T_{\text{eff}} = 124$ K, $g = 26$ m s^{-2}, and $\kappa \approx 5 \times 10^{-3}(P/1\text{ bar})$ $m^2 kg^{-1}$. This implies $P_0 \approx 0.2$ bar (20 000 Pa), which, given the simplicity of the calculation, is surprisingly close to the location of Jupiter's real tropopause where $T \approx 110$ K. Actually, the properties of the opacities of important absorbing chemical species like water and their pressure dependence imply that photospheres around 0.1 bar should be common (Robinson and Catling, 2014).

However, the Eddington boundary condition should not be used in the case of irradiated atmospheres because it does not properly account for both the incoming flux (mostly at visible wavelengths for planets around solar-type stars) and the intrinsic flux (in the infrared). The fact that opacities differ at these wavelengths yields the possibility of thermal inversions (higher visible than infrared opacities) or a greenhouse effect (lower visible than infrared opacities) and thus a hotter interior, something that cannot be captured without accounting for the different fluxes. Analytical solutions of the radiative transfer problem exist in the semi-gray case (two opacities for the visible and infrared, respectively) (Guillot, 2010; Hansen, 2008) and can even be extended to include non-gray effects (Parmentier and Guillot, 2014). Numerical solutions in the nonirradiated, solar composition case are provided by Saumon et al. (1996) and for the irradiation levels and compositions relevant for the solar system giant planets by Fortney et al. (2011). In that case, a grid is used to relate the atmospheric temperature and pressure at a given level to the radius R, intrinsic luminosity L, and incoming stellar luminosity $L_{*\text{p}}$: $r = R$; ($T_0 = T_0(R, L, L_{*\text{p}})$, $P_0 = P_0(R, L, L_{*\text{p}})$). P_0 is chosen to satisfy the condition that the corresponding optical depth at that level should be much larger than unity.

10.16.3.2 High-Pressure Physics and Equations of State

10.16.3.2.1 Hydrogen

In terms of pressures and temperatures, the interiors of the giant planets lie in a region for which accurate equations of state (EOS) are extremely difficult to calculate. This is because molecules, atoms, and ions can all coexist, in a fluid that is partially degenerate (free electrons have energies that are determined by both quantum and thermal effects) and partially coupled (Coulomb interactions between ions are not dominant but must be taken into account). The presence of many elements and their possible interactions further complicate matters. For lack of space, this section will mostly focus on hydrogen whose EOS has seen the most important developments in recent years. A phase diagram of hydrogen (Figure 5) illustrates some of the important phenomena that occur in the giant planets.

The photospheres of the giant planets are generally relatively cold (50–3000 K) and at low pressure (0.1–10 bar, or 10^4–10^6 Pa), so that hydrogen is in molecular form and the perfect gas conditions apply. As one goes deeper into the interior, hydrogen and helium progressively become fluid. (The perfect gas relation tends to underestimate the pressure by 10% or more when the density becomes larger than about 0.02 g cm^{-3} ($P \gtrsim 1$ kbar in the case of Jupiter).)

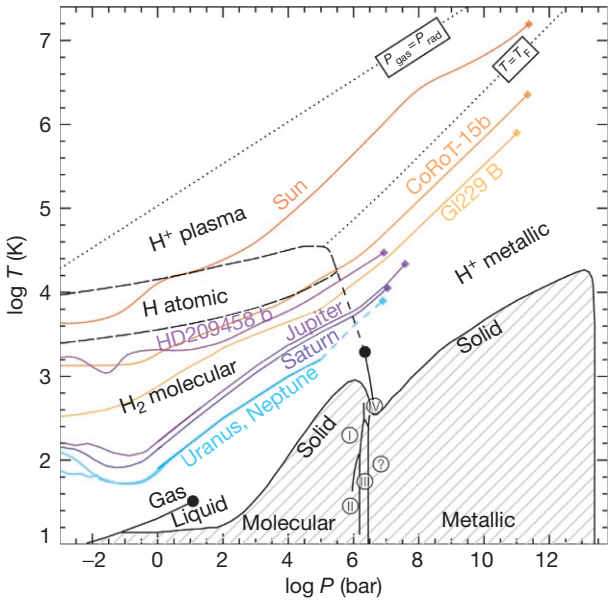


Figure 5 Phase diagram for hydrogen in the pressure–temperature plane, with pressure in bars (1 bar = 10^5 Pa = 10^6 dyn cm^{-2}). The thick lines indicate the first-order (discontinuous) phase transitions, the black circles, the critical points (McMahon et al., 2012; Morales et al., 2013b). Phase transitions in solid hydrogen with the different known phases are labeled I, II, III, and IV (McMahon et al., 2012). The region where 50% of all hydrogen is in atomic form (from Saumon et al., 1995) is shown by a thin contour. The approximate location of the molecular to metallic hydrogen (continuous) transition (Loubeyre et al., 2012; Morales et al., 2013b) is indicated by a dashed line. The $T = T_F$ line separates the nondegenerate region (at low pressures and high temperatures) from the region in which electrons are degenerate (see text). The $P_{gas} = P_{rad}$ line shows the region that is dominated by radiation pressure (at high temperatures). Colored lines show profiles for a selection of noteworthy substellar objects: giant planets from Uranus to Jupiter (Guillot, 2005), the hot Jupiter HD 209458 b (Burrows, 2013; Guillot and Showman, 2002), the brown dwarfs Gl 229 B (Marley et al., 1996), CoRoT-15 b (Bouchy et al., 2011), and our Sun (Christensen-Dalsgaard et al., 1996). The diamonds correspond to the conditions either at the core/envelope interface for Jupiter and Saturn or at the center for the other objects.

Characteristic interior pressures are considerably larger however: as implied by eqns [4] and [6], $P_c \approx GM^2/R^4$, of the order of 10–100 Mbar for Jupiter and Saturn. As shown in Figure 5, all the central pressures and temperatures of the giant planets and brown dwarfs (from Uranus to CoRoT-15b) lie in a regime of high pressures and temperatures lower than the corresponding Fermi temperature T_F, implying that electrons are degenerate: their pressure is mostly a function of the density of the material. In Jupiter and Saturn, the degeneracy parameter $\theta = T/T_F$ is always close to 0.03. Even for the warmer CoRoT-15b ($a \sim 60$ M_{Jup} brown dwarf discovered in transit in front of its parent star – see Bouchy et al., 2011), $\theta \approx 0.2$. This implies that for these objects, the thermal component is small so that the energy of electrons in the interior is expected to be only slightly larger than their nonrelativistic, fully degenerate limit: $u_e \geq 3/5\ kT_F = 15.6\ (\rho/\mu_e)^{2/3}$ eV, where k is Boltzmann's constant, μ_e is the number of nucleons per electron, and ρ is the density in g cm^{-3}. For pure hydrogen, when the density reaches $\sim$0.8 g cm^{-3}, the average energy of electrons becomes larger than hydrogen's ionization potential, even at zero temperature: hydrogen pressure ionizes and becomes metallic. This molecular to metallic transition occurs near Mbar pressures, but exactly how this happens is a result of the complex interplay of thermal, Coulomb, and degeneracy effects.

Recent laboratory measurements on fluid deuterium have been able to reach extremely high pressures up to 20 Mbar (Mochalov et al., 2012). Beyond that experimental feat, most of the progress of the decade in the domain has been the improvement in our understanding of the hydrogen metallization region at pressures of a fraction to a few Mbars, both from an experimental and numerical point of view (see the very complete review by McMahon et al., 2012). Already in the 1990s, gas-gun experiments had been able to measure a rise in the conductivity of molecular hydrogen up to $T \sim 3000$ K, $P \sim 1.4$ Mbar, a sign that metallization had been reached (Weir et al., 1996). A very sharp transition, probably discontinuous, was then later measured by isentropic convergent explosive shock experiments (Fortov et al., 2007) at pressures between 1.5 and 2.5 Mbar but uncertain temperatures below 4000 K (see McMahon et al., 2012). New experiments at higher temperatures using laser compression, directly (Sano et al., 2011) and from precompressed targets (Loubeyre et al., 2012), confirmed that this transition from a weakly conducting molecular fluid to a metal-like hydrogen fluid occurs around pressures near 1 Mbar and temperatures as high as 15 000 K, but that it is continuous at these temperatures. In parallel, ab initio calculations of the behavior of fluid hydrogen in this thermodynamical regime predicted the existence of a first-order liquid–liquid phase transition (the so-called PPT for *plasma phase transition*) with a critical point near $T \sim 1500$–2000 K and $P \sim 2.2$ Mbar (Morales et al., 2010, 2013b). This discontinuous transition at low temperatures merges into a continuous transition at temperatures above the critical point, in good agreement with the experimental data. The PPT is indicated by

a thick almost vertical line in Figure 5. It is prolonged by a dashed line indicating the location of the continuous transition from molecular to metallic hydrogen that extends up to the region of thermal dissociation and ionization of hydrogen.

The controversy that had arisen between laser-induced shock compression (Collins et al., 1998; da Silva et al., 1997) and pulsed-power shock compression (Knudson et al., 2004) regarding the maximum compression of deuterium along the principal shock Hugoniot has now been resolved in favor of the latter, thanks to new experiments (Boriskov et al., 2005; Hicks et al., 2009) and the realization that the equation of state of quartz used to calibrate the laser-induced shock experiments was incorrect (Knudson and Desjarlais, 2009). Similarly, the existence of a PPT of hydrogen at high temperatures in a regime crossing the adiabats of Jupiter and Saturn (Saumon et al., 1995) has now been shown to be a spurious effect resulting from the different treatment of molecules, atom, and ions within the so-called chemical picture (Chabrier et al., 2007). Both laboratory experiments and independent models based on first principles (Bonev et al., 2004; Desjarlais, 2003; French et al., 2012; Militzer et al., 2001; Vorberger et al., 2007) now agree and show that the transition from molecular to metallic hydrogen should occur continuously in all known giant planets.

Progress has also been made on the issue of the solidification of hydrogen (see McMahon et al., 2012, and references therein). This has led to the confirmation that the interiors of the hydrogen–helium giant planets and brown dwarfs are *fluid* whatever their age, a result expected since the pioneering study by Hubbard (1968). Of course, because of their initial gravitational energy, these objects are warm enough to avoid the critical point for the liquid–gas transition in hydrogen and helium, at very low temperatures, but they also lie comfortably above the solidification lines for hydrogen and helium. (An *isolated* Jupiter should begin partial solidification only after at least $\sim 10^3$ Ga of evolution.) They are considered to be fluid because at the high pressures and relatively modest temperatures in their interiors, Coulomb interactions between ions play an important role in the EOS and yield a behavior that is more reminiscent of that of a liquid than that of a gas, contrary to what is the case in, for example, solar-like stars. For Uranus and Neptune, the situation is actually more complex because at large pressures, they are not expected to contain a significant amount of hydrogen (see next section).

As Figure 5 highlights, while some highly irradiated planets and brown dwarfs like CoRoT-15b have temperature profiles that get close to the hydrogen thermal dissociation line, most of them are well within the molecular hydrogen regime at low pressures and in the metallic, degenerate regime at high pressures. Stars like our Sun lie in a higher temperature regime for which the EOS is dominated by thermal effects and electrons are essentially nondegenerate.

10.16.3.2.2 Other elements and mixtures

Hydrogen is of course a key element, but it is not sufficient to describe the structure of all giant planets. A description of the high-pressure behavior of other elements would go beyond the scope of the present review. We only sketch a few important results here.

In order to obtain tractable EOS in the entire domain of pressure and temperature spanned by the planets during their evolution, one has to consider simplifications, among which the first one is to consider that an element (e.g., hydrogen) dominates and that others can be considered as a perturbation. This is done for the hydrogen–helium mixture for the now classical EOS by Saumon et al. (1995) and now with more up-to-date EOSs (Caillabet et al., 2011; Militzer and Hubbard, 2013; Nettelmann et al., 2012). The addition of other elements can be done through the so-called additive volume rule that is generally a good approximation given other sources of uncertainty (Vorberger et al., 2007; see also Chabrier and Ashcroft, 1990).

EOS for elements other than hydrogen and helium in the parameter range relevant for giant planetary interiors have traditionally been difficult to obtain and are often not easily shared. Beyond an extrapolation from the classical ANEOS and Sesame tables (see Saumon and Guillot, 2004, and references therein), new results have become available. In particular, ab initio simulations revealed that when compressing water or ammonia along an isentrope from conditions relevant to the atmospheres of Uranus and Neptune, they transition from a molecular to an ionic fluid, then to a superionic fluid, and finally to a plasma (Cavazzoni et al., 1999). The superionic fluid corresponds to a state in which protons move relatively freely among a lattice of oxygen atoms. An equation of state for water has been calculated from first principles by French et al. (2009) and is found to be in good agreement with experiments (Knudson et al., 2012). A similar equation of state for ammonia is presented by Bethkenhagen et al. (2013). In models of Uranus and Neptune, for water, the ionic transition occurs near 0.1 Mbar and the superionic transition near 1 Mbar (Redmer et al., 2011).

In some cases, however, generally at low enough temperatures for a given pressure range, mixtures cannot remain homogeneous. This further complicates the calculation of the EOS and has important physical consequences for the planetary structure: the two components having different molecular weights, they tend to be separated by gravity so that the heavier component settles down under the lighter one. This is the case of the hydrogen and helium mixture for which it was proposed that such a separation would occur in Saturn already in the 1970s (Salpeter, 1973; Stevenson and Salpeter, 1977a), but for which realistic calculations have only become possible in the past decade or so (see Lorenzen et al., 2011; Morales et al., 2009, 2013a). Figure 6 shows the comparison between two of these calculations based on first-principles simulations. According to these calculations, Saturn's interior is in the phase separation region below 1 Mbar but whether Jupiter is too depends on which calculation is considered.

The separation of other mixtures has also been calculated and is relevant to understand the initial formation and subsequent possible erosion of the cores of the giant planets (see Guillot et al., 2004). First-principles calculations of the water in metallic hydrogen (at pressures above 10 Mbar) predict a critical phase separation temperature of less than 4000 K (Wilson and Militzer, 2010), implying that water is completely soluble in hydrogen in the interiors of Jupiter, Saturn, and generally gas giants. This is also the case of iron, with a critical

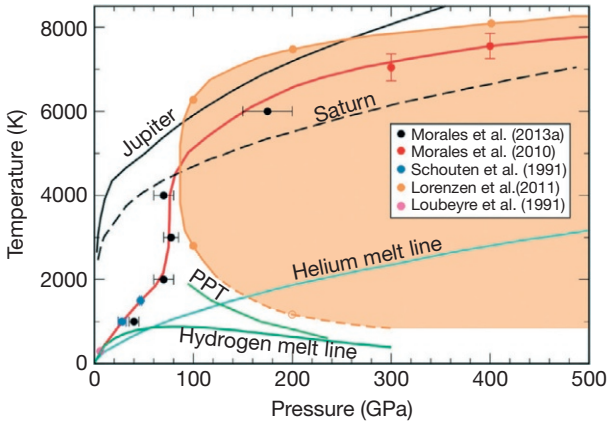


Figure 6 Phase diagram for the hydrogen–helium mixture for a helium mole concentration of 8%. The orange region shows where the two elements are expected to separate from each other according to the calculations of Lorenzen et al. (2011). The red curve corresponds to the critical temperature for that separation according to Morales et al. (2013a). Numerical results by Schouten et al. (1991) and experimental determinations by Loubeyre et al. (1991) are also shown. The back curves show the isentropes of Jupiter (plain) and Saturn (dashed), respectively. Reproduced from Morales MA, Hamel S, Caspersen K, and Schwegler E (2013) Hydrogen-helium demixing from first principles: From diamond anvil cells to planetary interiors. *Physical Review B* 87(17): 174105.

temperature of around 2000 K (Wahl et al., 2013). Finally, mixing rocks (specifically MgO) and metallic hydrogen is also relatively easy, even though the critical temperature for the same pressure range is higher, of order 10 000 K or less (Wilson and Militzer, 2012).

10.16.3.3 Heat Transport

The giant planets possess hot interiors, implying that a relatively large amount of energy has to be transported from the deep regions of the planets to their surface. This can be done by either radiation, conduction, or, if these processes are not sufficient, convection. Convection is generally ensured by the rapid rise of the opacity with increasing pressure and temperature. At pressures of a bar or more and relatively low temperatures (<1000 K), the three dominant sources of opacities are water, methane, and collision-induced absorption by hydrogen molecules.

However, in the intermediate temperature range between $\sim$1200 and 1500 K, the Rosseland opacity due to the hydrogen and helium absorption behaves differently: the absorption at any given wavelength increases with density, but because the temperature also rises, the photons are emitted at shorter wavelengths, where the monochromatic absorption is smaller. As a consequence, the opacity can decrease. This was shown by Guillot et al. (2004) to potentially lead to the presence of a deep radiative zone in the interiors of Jupiter, Saturn, and Uranus.

This problem must however be reanalyzed in the light of observations and analyses of brown dwarfs. Their spectra show unexpectedly wide sodium and potassium absorption lines (see Burrows et al., 2000), in spectral regions where hydrogen, helium, water, methane, and ammonia are relatively transparent. It thus appears that the added contribution of these elements (if they are indeed present) would wipe out any radiative region at these levels (Guillot et al., 2004).

At temperatures above 1500–2000 K, two important sources of opacity appear: (i) the rising number of electrons greatly enhances the absorption of H_2^- and H^-; (ii) TiO, a very strong absorber at visible wavelengths, is freed by the vaporization of $CaTiO_3$. Again, the opacity rises rapidly that ensures a convective transport of the heat. Still deeper, conduction by free electrons becomes more efficient, but the densities are found not to be high enough for this process to be significant, except perhaps near the central core (see Hubbard, 1968; Stevenson and Salpeter, 1977b).

While our giant planets seem to possess globally convective interiors, strongly irradiated extrasolar planets must develop a radiative zone just beneath the levels where most of the stellar irradiation is absorbed. Depending on the irradiation and characteristics of the planet, this zone may extend down to kbar levels, the deeper levels being convective. In this case, a careful determination of the opacities is necessary (but generally not possible) as these control the cooling and contraction of the deeper interior (see Freedman et al., 2008, for a discussion of opacities and tables for substellar atmospheres and interiors).

10.16.3.4 The Contraction and Cooling Histories of the Giant Planets

The interiors of the giant planets are expected to evolve with time from a high entropy, high θ value, hot initial state to a low entropy, low θ, cold degenerate state. The essential underlying physics can be derived from the well-known virial theorem and the energy conservation that link the planet's internal energy E_i, gravitational energy E_g, and luminosity through:

$$\xi E_i + E_g = 0 \quad [8]$$

$$L = -\frac{\xi - 1}{\xi}\frac{dE_g}{dt} \quad [9]$$

where $\xi = \int_0^M 3(P/\rho)dm \int_0^M udm \approx \langle 3P/\rho u \rangle$ the brackets indicating averaging, and u is the specific internal energy. For a diatomic perfect gas, $\xi = 3.2$; for fully degenerate nonrelativistic electrons, $\xi = 2$.

Thus, for a giant planet or brown dwarf beginning its life mostly as a perfect H_2 gas, two-thirds of the energy gained by contraction is radiated away, one-third being used to increase E_i. The internal energy being proportional to the temperature, the effect is to heat up the planet. This represents the slightly counterintuitive but well-known effect that a star or giant planet initially heats up while radiating a significant luminosity (e.g., Kippenhahn and Weigert, 1994).

Let us now move further in the evolution, when the contraction has proceeded to a point where the electrons have become degenerate. For simplicity, we will ignore Coulomb interactions and exchange terms, and assume that the internal energy can be written as $E_i = E_{el} + E_{ion}$ and that furthermore $E_{el} \gg E_{ion}$ (θ is small). Because $\xi \approx 2$, we know that half of the gravitational potential energy is radiated away and half of it

goes into internal energy. The problem is to decide how this energy is split into an electronic and an ionic part. The gravitational energy changes with some average value of the interior density as $E_g \propto 1/R \propto \rho^{1/3}$. The energy of the degenerate electrons is essentially the Fermi energy: $E_{el} \propto \rho^{2/3}$. Therefore, $\dot{E}_g \approx 2(E_g/E_{el})\dot{E}_{el}$. Using the virial theorem and specifically eqn [8], we get that $\dot{E}_g \approx -\dot{E}_{el}$. The luminosity is by definition $L = -(\dot{E}_g + \dot{E}_i)$ and therefore

$$L \approx -\dot{E}_{ion} \propto -\dot{T} \qquad [10]$$

In this limit, the gravitational energy lost is entirely absorbed by the increase in pressure of the degenerate electrons and the observed luminosity is due to the thermal cooling of the ions. The luminosity can be written in terms of a modified Kelvin–Helmholtz formula:

$$L \approx \eta \frac{GM^2}{R\tau} \qquad [11]$$

where τ is the age and η is a factor that hides most of the complex physics. In the approximation that Coulomb and exchange terms can be neglected, $\eta \approx \theta/(\theta+1)$. The poor compressibility of the giant planets in their mature evolution stages implies that $\eta \ll 1$ ($\eta \sim 0.03$ for Jupiter): the luminosity is not obtained from the entire gravitational potential but from the much more limited reservoir constituted by the thermal internal energy. Equation [11] shows that to first order, $\log L \propto -\log \tau$: very little time is spent at high luminosity values. In other words, the problem is (in most cases) weakly sensitive to initial conditions. However, it is to be noticed that with progress in our capability to detect very young objects, that is, planets and brown dwarfs of only a few million years of age, the problem of the initial conditions does become important (Marley et al., 2007). Interestingly, at these early stages, their luminosity appears to strongly depend on their core mass (Mordasini, 2013).

Figure 7 shows more generally how not only the giant planets but also brown dwarfs and small stars see their luminosities evolve as a function of time. The $1/\tau$ slope is globally conserved, with some variations for brown dwarfs, during the transient epoch of deuterium burning, and of course for stars, when they begin burning efficiently their hydrogen and settle

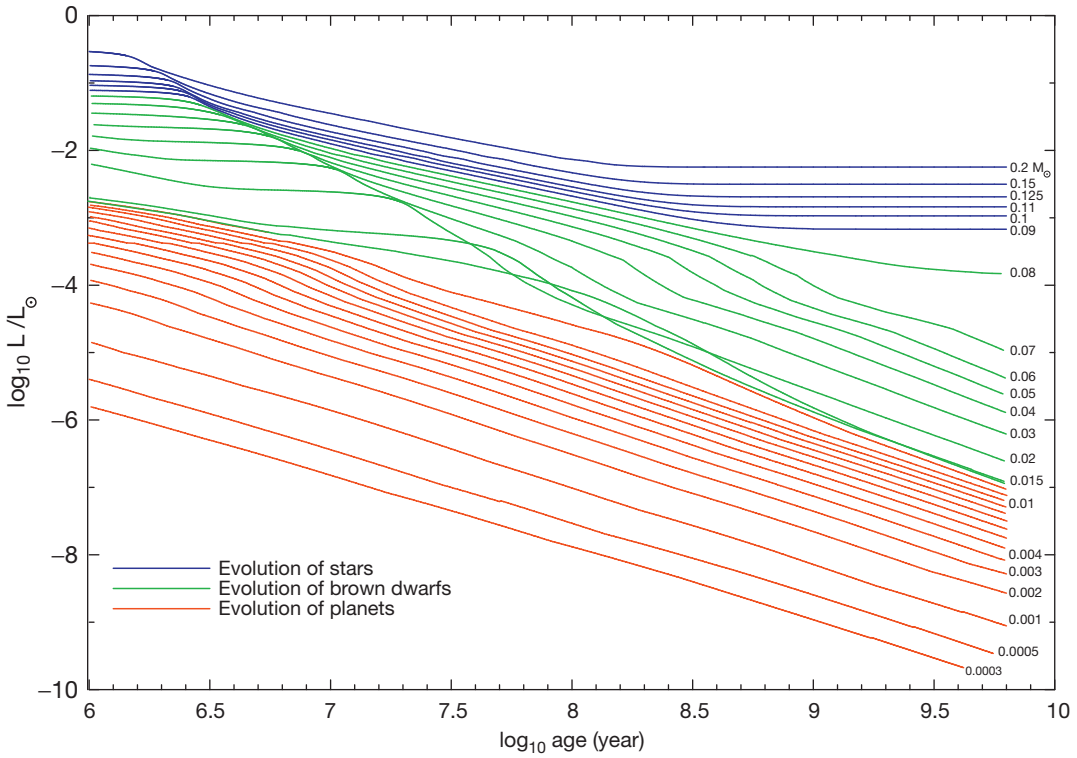


Figure 7 Evolution of the luminosity (in $L_\odot$) of solar-metallicity M dwarfs and substellar objects versus time (in year) after formation. In this figure, 'brown dwarfs' are arbitrarily designated as those objects that burn deuterium, while those that do not are tentatively labeled 'planets.' Stars are objects massive enough to halt their contraction due to hydrogen fusion. Each curve is labeled by its corresponding mass in $M_\odot$, with the lowest three corresponding to the mass of Saturn, half the mass of Jupiter, and the mass of Jupiter. Reproduced from Burrows A, Marley M, Hubbard WB, et al. (1997) A nongray theory of extrasolar giant planets and brown dwarfs. *Astrophysical Journal* 491: 856.

on the main sequence: in that case, the tendency of the star to contract under the action of gravity is exactly balanced by thermonuclear hydrogen fusion.

10.16.3.5 Mass–Radius Relation

The relation between mass and radius has very fundamental astrophysical applications. Most importantly, it allows one to infer the gross composition of an object from a measurement of its mass and radius. This is especially relevant in the context of the discovery of extrasolar planets with both radial velocimetry and the transit method, as the two techniques yield relatively accurate determination of M and R, these determinations being often limited by the uncertainty on the stellar parameters themselves.

Figure 8 shows mass–radius relations for compact degenerate objects from the giant planets to brown dwarfs and low-mass stars. The right-hand side of the diagram shows a rapid increase of the radius with mass in the stellar regime that is directly due to the onset of stable thermonuclear reactions. In this regime, observations and theoretical models agree (see however Ribas, 2006, for a more detailed discussion). The left-hand side of the diagram is obviously more complex, and this can be understood by the fact that planets have much larger variations in composition than stars and because external factors such as the amount of irradiation they receive do affect their contraction in a significant manner.

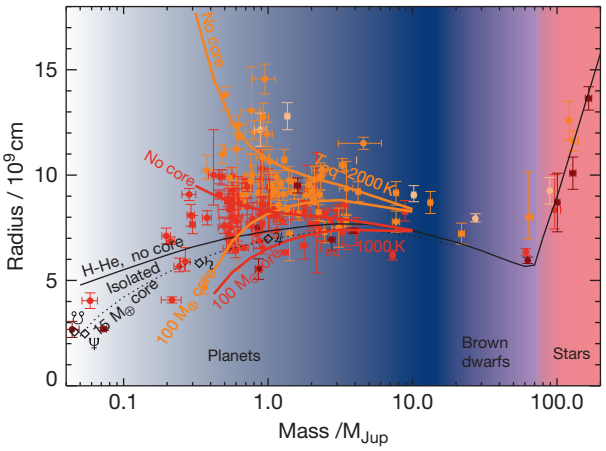


Figure 8 Theoretical and observed mass–radius relations. The black line is applicable to the evolution of solar composition planets, brown dwarfs, and stars, when isolated or nearly isolated (as Jupiter, Saturn, Uranus, and Neptune, defined by diamonds and their respective symbols), after 5 Ga of evolution. The dotted line shows the effect of a 15 $M_\oplus$ core on the mass–radius relation. Orange and yellow curves represent the mass–radius relations for heavily irradiated planets with equilibrium temperatures of 1000 and 2000 K, respectively, and assuming that 0.5% of the incoming stellar luminosity is dissipated at the center (see **Section 10.16.4.3**). For each irradiation level, two cases are considered: a solar composition planet with no core (top curve) and one with a 100 $M_\oplus$ central core (bottom curve). Circles with error bars correspond to known planets, brown dwarfs, and low-mass stars, color-coded as a function of their equilibrium temperature (below 750, 1500, 2250 K, and above 2250 K, respectively, from darkest to lightest).

Let us first concentrate on isolated or nearly isolated gaseous planets. The black curves have a local maximum near 4 M_J: At small masses, the compression is small so that the radius increases with mass. At large masses, degeneracy sets in and the radius decreases with mass.

This can be understood on the basis of polytropic models based on the assumption that $P = K\rho^{1+1/n}$, where K and n are constants. Because of degeneracy, a planet of large mass will tend to have $n \to 1.5$, while a planet of smaller mass will be less compressible ($n \to 0$). Indeed, it can be shown that in their inner 70–80% in radius, isolated solar composition planets of 10, 1, and 0.1 M_J have $n = 1.3$, 1.0, and 0.6, respectively. From polytropic equations (e.g., Chandrasekhar, 1939):

$$R \propto K^{n/(3-n)} M^{(1-n)/(3-n)} \qquad [12]$$

Assuming that K is independent of mass, one gets $R \propto M^{0.16}$, M^0, and M^{-018} for $M = 10$, 1, and 0.1 M_J, respectively, in relatively good agreement with Figure 8 (the small discrepancies are due to the fact that the intrinsic luminosity and hence K depend on the mass considered).

Figure 8 shows already that the planets in our solar system are not made of pure hydrogen and helium and require an additional fraction of heavy elements in their interior, either in the form of a core or distributed in the envelope (dotted line).

For extrasolar planets, the situation is complicated by the fact that the intense irradiation that they receive plays a major role in their evolution. The present sample is already quite diverse, with equilibrium temperature (defined as the effective temperature corresponding to the stellar flux received by the planet) ranging from 1000 to 2500 K. Their compositions are also quite variable, with some planets having large masses of heavy elements (Guillot et al., 2006; Sato et al., 2005). The orange and yellow curves in Figure 8 show theoretical results for equilibrium temperatures of 1000 and 2000 K, respectively. Two extreme models have been plotted: assuming a purely solar composition planet (top curve) and assuming the presence of a 100 $M_\oplus$ central core (bottom curve). In each case, an additional energy source proportional to 0.5% of the incoming luminosity was also assumed (see discussion in **Section 10.16.4.3** hereafter).

The increase in radius for decreasing planetary mass for irradiated, solar composition planets with little or no core can be understood using the polytropic relation (eqn [12]) but accounting for variations of K as defined by the atmospheric boundary condition. Using the Eddington approximation, assuming $\kappa \propto P$ and a perfect gas relation in the atmosphere, one can show that $K \propto (M/R^2)^{-1/2n}$ and that therefore $R \propto M^{(1/2-n)/(2-n)}$. With $n = 1$, one finds $R \propto M^{-1/2}$. Strongly irradiated hydrogen–helium planets of small masses are hence expected to have the largest radii that qualitatively explain the positions of the extrasolar planets in Figure 8. Note that this estimate implicitly assumes that n is constant throughout the planet. The real situation is more complex because of the growth of a deep radiative region in most irradiated planets and because of structural changes between the degenerate interior and the perfect gas atmosphere (Guillot, 2005).

In the case of the presence of a fixed mass of heavy elements, the trend is inverse because of the increase of mean

molecular mass (or equivalently core/envelope mass) with decreasing total mass. Thus, small planets with a core are much more tightly bound and less subject to evaporation than those that have no core.

10.16.3.6 Rotation and the Figures of Planets

The mass and radius of a planet informs us on its global composition. Because planets are also rotating, one is allowed to obtain more information on their deep interior structure. The hydrostatic equation becomes more complex however:

$$\frac{\nabla P}{\rho} = \nabla\left(G\iiint\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d^3\mathbf{r}'\right) - \boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r}) \qquad [13]$$

where $\boldsymbol{\Omega}$ is the rotation vector. The resolution of eqn [13] is a complex problem. It can however be somewhat simplified by assuming that $|\boldsymbol{\Omega}| = \omega$ is such that the centrifugal force can be derived from a potential. The hydrostatic equilibrium then writes $\nabla P = \rho\nabla U$, and the *figure* of the rotating planet is then defined by the $U = constant$ level surface.

One can show (e.g., Zharkov and Trubitsyn, 1978) that the hydrostatic equation of a fluid planet can then be written in terms of the mean radius $\bar{r}$ (the radius of a sphere containing the same volume as that enclosed by the considered equipotential surface):

$$\frac{1}{\rho}\frac{\partial P}{\partial \bar{r}} = -\frac{Gm}{\bar{r}^2} + \frac{2}{3}\omega^2\bar{r} + \frac{GM}{\bar{R}^3}\bar{r}\varphi_\omega \qquad [14]$$

where M and $\bar{R}$ are the total mass and mean radius of the planet and φ_ω is a slowly varying function of $\bar{r}$. (In the case of Jupiter, φ_ω varies from about 2×10^{-3} at the center to 4×10^{-3} at the surface.) Equations [5]–[7] remain the same with the hypothesis that the level surfaces for the pressure, temperature, and luminosity are equipotentials. The significance of rotation is measured by the ratio of the centrifugal acceleration to the gravity:

$$q = \frac{\omega^2 R_{eq}^3}{GM} \qquad [15]$$

As discussed in **Section 10.16.2.2**, in some cases, the external gravity field of a planet can be accurately measured in the form of gravitational moments J_k (with zero odd moments for a planet in hydrostatic equilibrium) that measure the departure from spherical symmetry. Together with the mass, this provides a constraint on the interior density profile (see Zharkov and Trubitsyn, 1974; see also **Chapters 10.02** and **10.04**):

$$M = \iiint\rho(r,\theta)d^3\tau$$
$$J_{2i} = -\frac{1}{MR_{eq}^{2i}}\iiint\rho(r,\theta)r^{2i}P_{2i}(\cos\theta)d^3\tau$$

where $d\tau$ is a volume element and the integrals are performed over the entire volume of the planet.

Figure 9 shows how the different layers inside a planet contribute to the mass and the gravitational moments. The figure applies to Jupiter but would remain relatively similar for other planets. Note however that in the case of Uranus and Neptune, the core is a sizable fraction of the total planet and contributes to both J_2 and J_4. Measured gravitational moments

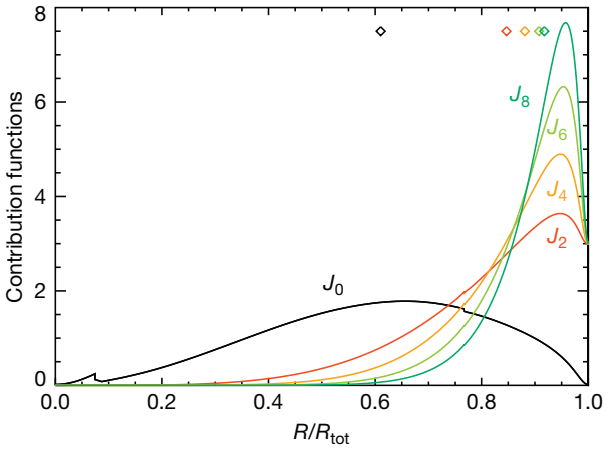


Figure 9 Contribution of the level radii to the gravitational moments of Jupiter. J_0 is equivalent to the planet's mass. The small discontinuities are caused by the following transitions, from left to right: core/envelope, helium-rich/helium-poor (metallic/molecular). Diamonds indicate the median radius for each moment. They correspond to pressures of 9.1, 1.3, 0.75, 0.45, and 0.35 Mbar, respectively, from left to right (J_0–J_8).

thus provide information on the external levels of a planet. It is only indirectly, through the constraints on the outer envelope that the presence of a central core can be inferred. As a consequence, it is impossible to determine this core's state (liquid or solid), structure (differentiated, partially mixed with the envelope), and composition (rock, ice, helium, etc.) from the gravity field data.

The Juno (Bolton, 2010) and Cassini Solstice (Spilker, 2012) missions are expected to yield considerable improvements in our determination of the gravity fields of Jupiter and Saturn, respectively. Because the theory of figures is limited by its expansion in terms of the rotation parameter q and because purely barotropic solutions are possible only in the limit of solid-body rotation and of pure rotation on cylinders, these high-precision measurements will require new approaches to include rotation in planetary models (se Hubbard, 1999, 2013). Separately, these measurements will enable new constraints such as the determination of the planets' angular momentum through the measurement of the relativistic Lense–Thirring effect (Iorio, 2010) and the determination of their moment of inertia (Helled, 2011; Helled et al., 2011b). But probably, the most important prospects lie in the possibility to couple measurements on gravity field, magnetic fields, and wind speeds with combined tridimensional magnetohydrodynamic models (see **Sections 10.16.2.3** and **10.16.2.4**).

For planets outside the solar system, although measuring their gravitational potential is presently beyond reach, an indirect measurement of the planets' Love number k_2 may be possible in systems of planets in which one is locked into a so-called fixed-point eccentricity. So far, one such transiting system is known, HAT-P-13 (Batygin et al., 2009; Mardling, 2010), because it contains a transiting planet, HAT-P-13b, and a companion, HAT-P-13c, whose minimum mass $M\sin i$ and orbital eccentricity are known (Bakos et al., 2009). The value of k_2 constrains the interior structure in a way that is very similar to J_2 and has been used to obtain first constraints on the interior structure of this planet (Kramm et al., 2012).

10.16.4 Interior Structures and Evolutions

10.16.4.1 Jupiter and Saturn

As illustrated by Figure 10, the simplest interior models of Jupiter and Saturn matching all observational constraints assume the presence of three main layers: (i) an outer hydrogen–helium envelope, whose global composition is that of the deep atmosphere; (ii) an inner hydrogen–helium envelope, enriched in helium because the whole planet has to fit the H/He protosolar value; and (iii) a central dense core. Because the planets are believed to be mostly convective, these regions are expected to be globally homogeneous. (Many interesting thermochemical transformations take place in the deep atmosphere, but they are of little concern to us.)

The transition from a helium-poor upper envelope to a helium-rich lower envelope is thought to take place through the formation of helium-rich droplets that fall deeper into the planet due to their larger density. These droplets form when the temperature–pressure profiles enter the separation region for the initial helium abundance, as shown in Section 10.16.3.2.2. Three-layer models implicitly make the hypothesis that this region is adiabatic (this is justified only if convection is not inhibited by the formation of helium droplets, as discussed by Stevenson and Salpeter (1977a)) and narrow. Figure 6 shows that this zone may be extended, especially in present-day Saturn (see also Fortney and Hubbard, 2003).

As discussed by Stevenson and Salpeter (1977a) and Stevenson (1982), the planets would start from an initially hot and homogeneous state and would start entering the phase separation region progressively, leading to a depletion of helium in the outer region and its increase in the deeper interior. According to the calculations by Lorenzen et al. (2011) and Morales et al. (2013a), the separation would first occur at a pressure between 1 and 2 Mbar and the inhomogeneous region would grow inward but not so much toward lower pressures because of the higher solubility of helium in molecular hydrogen. According to the simulations, although needed to explain the abundances in Jupiter's atmosphere (Section 10.16.2.6), it is not yet clear that the process has begun in this planet. Fully consistent calculations that account both for the constraints on the planets' gravitational moments and their atmospheric composition should become possible, especially with the additional information brought by the Juno (Bolton, 2010) and Cassini Solstice (Spilker, 2012) space missions.

In the absence of these calculations, adiabatic three-layer models can be used as a useful guidance to a necessarily hypothetical ensemble of allowed structures and compositions of Jupiter and Saturn. A relatively extensive exploration of the parameter space has been performed by several authors (Fortney and Nettelmann, 2010; Helled and Guillot, 2013; Nettelmann et al., 2012, 2013b; Saumon and Guillot, 2004). The calculations account for a transition from a helium-poor outer envelope to a helium-rich inner envelope. The abundance of heavy elements may or may not be held constant across this transition. Many sources of uncertainties are taken into account however; among them, the most significant are on the EOS of hydrogen and helium, the uncertain values of J_4 and J_6, the presence of differential rotation deep inside the planet, the location of the helium-poor to helium-rich region, and the uncertain helium to hydrogen protosolar ratio.

Their results indicate that Jupiter's core is smaller than $\sim 10\ M_\oplus$ and that its global composition is pretty much unknown (between 10 and 42 $M_\oplus$ of heavy elements in total). The models indicate that Jupiter is enriched compared to the solar value by a factor 1.5–8 times the solar value. This enrichment is compatible with a global uniform enrichment of all species near the atmospheric Galileo values but allows many other possibilities as well.

Other models of Jupiter based on an ab initio equation of state by Militzer et al. (2008) led to a solution with a large core mass and a very small enrichment in heavy elements in the

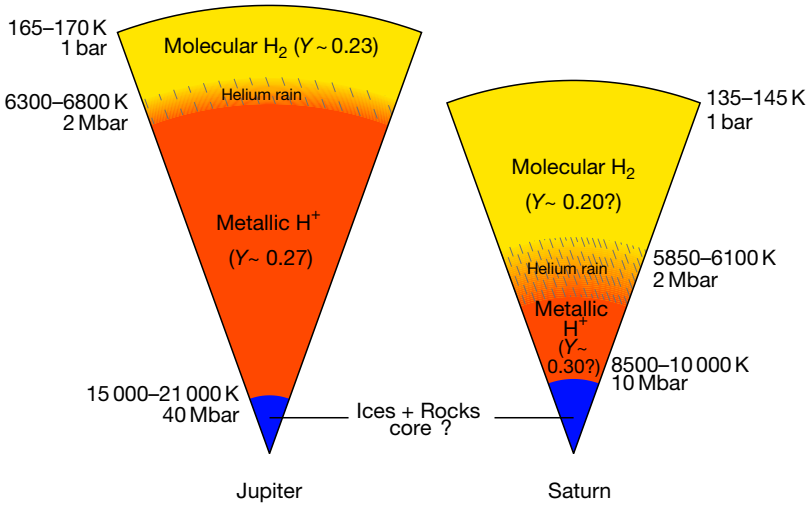


Figure 10 Schematic representation of the interiors of Jupiter and Saturn. The range of temperatures is estimated using homogeneous models and including a possible radiative zone indicated by the hashed regions. Helium mass mixing ratios Y are indicated. The size of the central rock and ice cores of Jupiter and Saturn is very uncertain (see text). In the case of Saturn, the inhomogeneous region may extend down all the way to the core, which would imply the formation of a helium core. Adapted from Guillot T (1999) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77.

envelope incompatible with either the Galileo probe measurements or the protosolar helium abundance. A significant update in the EOS is presented by Militzer and Hubbard (2013). This updated EOS yields a warmer interior than the 2008 models and should therefore lead to a smaller core mass and larger amount of heavy elements in the envelope, in line with the other results. At the same time, the Militzer and Hubbard (2013) EOS is much more accurate than the range of EOSs used by Saumon and Guillot (2004), and it differs slightly from the other ab initio EOS used by Nettelmann et al. (2012) and Nettelmann et al. (2013b). New constraints should therefore be derived on the basis of those new data.

In the case of Saturn, the solutions depend less on the hydrogen EOS because the Mbar pressure region is comparatively smaller. The total amount of heavy elements present in the planet can therefore be estimated with a better accuracy than for Jupiter and is between 16 and 30 $M_\oplus$ (Helled and Guillot, 2013; Nettelmann et al., 2013b). The uncertainty on the core mass is found to be larger than for Jupiter because a heavy element-rich inner envelope can mimic the gravitational signature of the core. As a result, only an upper limit on the core mass of 20 $M_\oplus$ is derived.

Concerning the *evolutions* of Jupiter and Saturn, the three main sources of uncertainty are, by order of importance, (1) the magnitude of the helium separation, (2) the EOS, and (3) the atmospheric boundary conditions. **Figure 11** shows an ensemble of possibilities that attempts to bracket the minimum and maximum cooling. In all these quasi-adiabatic cases, helium sedimentation is needed to explain Saturn's present luminosity (see Fortney and Hubbard, 2003; Hubbard, 1977; Hubbard et al., 1999; Salpeter, 1973; Stevenson and Salpeter, 1977a). In the case of Jupiter, the sedimentation of helium that appears to be necessary to explain the low atmospheric helium abundance poses a problem for evolution models because it appears to generally prolong its evolution beyond 4.55 Ga, the age of the solar system (Fortney et al., 2011; Nettelmann et al., 2012). However, different solutions are possible, including improvements of the EOS and atmospheric boundary conditions or even the possible progressive erosion of the central core that would yield a lower luminosity of Jupiter at a given age (Guillot et al., 2004).

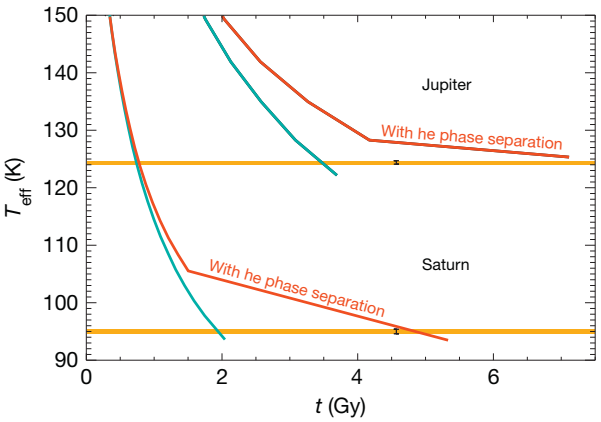


Figure 11 Final stages of evolution of Jupiter and Saturn. The present effective temperatures, reached after ~4.55 Ga of evolution, are indicated as horizontal orange lines. For each planet, two models represent attempts to bracket the ensemble of possibilities, with the faster evolution corresponding to that of a homogeneous planet, while the slowest evolution includes the effect of helium settling in the last evolution phase. Adapted from Hubbard WB (1999) NOTE: Gravitational signature of Jupiter's deep zonal flows. *Icarus* 137: 357–359; Fortney JJ and Hubbard WB (2003) Phase separation in giant planets: Inhomogeneous evolution of Saturn. *Icarus* 164: 228–243.

A different set of solutions appears when one considers the possibility that the envelopes of Jupiter and Saturn are *not* homogeneous and adiabatic on large scales, but are instead of variable composition, with heavier elements at the bottom. In that case, convection can be strongly inhibited that requires the temperature gradient to be larger in order to transport the same intrinsic luminosity. In the case of the giant planets, this process, known as semiconvection or diffusive convection, leads to the formation of a nonstatic staircase structure with diffusive interfaces with abrupt variations of temperature and composition and a homogeneous adiabatic structure in between (Rosenblum et al., 2011; Stevenson, 1985). By assuming that this structure is maintained over the entire envelopes of Jupiter and Saturn, Leconte and Chabrier (2012) derived much warmer interior structures with also 30–60% more heavy elements than in conventional models. However, the assumption that this inhomogeneous composition is maintained in the entire envelopes is ad hoc. In reality, one may expect semiconvection to be confined to a much smaller region and thus have a more limited effect. The problem is open however and requires further study. Its understanding is also critical for deciding whether Jupiter's core can erode into its envelope (see Guillot et al., 2004; Wilson and Militzer, 2012).

As discussed in **Sections 10.16.2.3** and **10.16.2.4**, classical interior models of Jupiter and Saturn based on the assumption of a homogeneous structure in the molecular envelope and an increase of the conductivity mainly due to hydrogen metallization do yield solutions for their magnetic fields and atmospheric zonal winds that globally match the observations. However, important 'details' such as why Saturn's magnetic field is axisymmetric and Jupiter's is not remain unexplained. Coupling interior and dynamical models in order to fit both Jupiter and Saturn's gravity and magnetic fields as well as their observed zonal winds (see **Sections 10.16.2.3** and **10.16.2.4**) should bring a more global understanding of the planetary structures.

10.16.4.2 Uranus and Neptune

Although the two planets are relatively similar, **Table 3** already shows that Neptune's larger mean density compared to Uranus has to be due to a slightly different composition: either more heavy elements compared to hydrogen and helium or a larger rock to ice ratio. The gravitational moments impose that the density profiles lie close to that of 'ices' (a mixture initially composed of, for example, H_2O, CH_4, and NH_3 but that rapidly becomes an ionic fluid of uncertain chemical composition in the planetary interior), except in the outermost layers, which have a density closer to that of hydrogen and helium (Marley et al., 1995; Podolak et al., 2000). As illustrated in **Figure 12**, three-layer models of Uranus and Neptune consisting of a central 'rocks' core (magnesium silicate and iron material), an ice layer, and a hydrogen–helium gas envelope have been

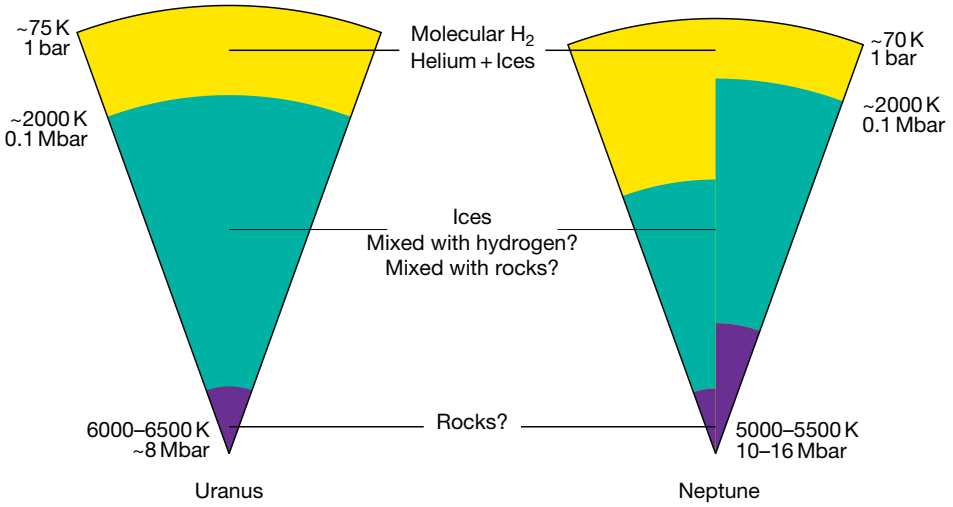


Figure 12 Schematic representation of the interiors of Uranus and Neptune. The ensemble of possibilities for Neptune is larger. Two possible structures are shown. Adapted from Guillot T (1999) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77 using results from Nettelmann N, Helled R, Fortney JJ, and Redmer R (2013) New indication for a dichotomy in the interior structure of Uranus and Neptune from the application of modified shape and rotation data. *Planetary and Space Science* 77: 143–151.

calculated (Fortney and Nettelmann, 2010; Helled et al., 2011a; Hubbard et al., 1995; Nettelmann et al., 2013a; Podolak et al., 1991).

According to the models of Nettelmann et al. (2013a), Uranus contains a minimum of 1.8–2.2 $M_\oplus$ of hydrogen and helium and Neptune 2.7–3.3 $M_\oplus$. The global ice to rock ratio that is derived is very high (19–36) in Uranus, while Neptune has a wide range of solutions from 3.6 to 14. These values are much larger than the canonical ice to rock ratio of 2–3 for the protosun that accounts for the abundances of all elements condensing at low temperatures ('ices') versus that of more refractory elements ('rocks'). The fact that either planet would have accreted much less rocks than ices is puzzling and unexplained by formation models. It is probably an artifact from assuming ices being confined to the envelope and rocks to the core.

The evolution of the two planets also remains a mystery. While Neptune's present luminosity may be explained by the adiabatic cooling of the planet over the age of the solar system, this is not the case of Uranus's very low luminosity (Fortney et al., 2011; Nettelmann et al., 2013a; Podolak et al., 1995). This could be explained by the presence of a strongly inhibiting compositional gradient decoupling an inner region that would remain hot and an outer envelope that would cool progressively (Podolak et al., 1995). Such regions could also be present in Neptune but considerably deeper. Unfortunately, this qualitative explanation cannot be tied to the inferred interior structures. Apart from the latest models by Nettelmann et al. (2013a), the models of Uranus and Neptune are too similar (and so are their magnetic fields – see **Section 10.16.2.3**) to explain why Uranus would have such a small intrinsic heat flux and not Neptune.

In fact, it is likely that all present models of Uranus and Neptune are inadequate because of the assumption of an adiabatic temperature structure across interfaces with different compositions. Instead, diffusive convection should occur and lead to strongly superadiabatic temperature gradient (e.g., Rosenblum et al., 2011). As in the case of Jupiter and Saturn (see Leconte and Chabrier, 2012), this would lead to higher temperatures in the interior and very different constraints on the interior composition. The amount of rocks required to fit the mean density and gravitational moments would certainly rise, potentially solving the ice to rock ratio problem. The evolution of the planets would be very different as the present-day luminosity would be mostly governed by the leak of heat from the hot interior by diffusion at the interfaces.

10.16.4.3 Irradiated Giant Planets

10.16.4.3.1 Interior structure and dynamics

The physics that governs the calculation of the interior structure and evolution models of the giant planets described in the previous sections can be applied in principle to any gaseous exoplanet and brown dwarf. We focus the discussion on the ones that orbit extremely close to their star because of the possibility to directly characterize them and measure their mass, radius, and in some cases even the properties of their atmosphere. Two planets are proxies for this new class of objects: the first extrasolar giant planet discovered, 51 Peg b, with an orbital period of $P = 4.23$ days, and the first *transiting* extrasolar giant planet, HD 209458 b, with $P = 3.52$ days. Following widespread usage, we call these planets 'hot Jupiters' (aka 'Pegasids' since these two archetypes have been discovered in the Pegasus constellation).

With such a short orbital period, these planets are for most of them subject to an irradiation from their central star that is so intense that the absorbed stellar energy flux can be about $\sim 10^4$ times larger than their intrinsic flux. The atmosphere is thus prevented from cooling, with the consequence that a radiative zone develops and governs the cooling and contraction of the interior (Guillot et al., 1996). Typically, for a planet like HD 209458 b, this radiative zone extends to kbar levels, $T \sim 4000$ K, and is located in the outer 5% in radius (0.3% in mass) (Guillot and Showman, 2002).

Problems in modeling the evolution of hot Jupiters arise because of the uncertain outer boundary condition. The intense stellar flux implies that the atmospheric temperature profile is extremely dependent upon the opacity sources considered. Depending on the chosen composition, the opacity data used, the assumed presence of clouds, and the geometry considered, resulting temperatures in the deep atmosphere can differ by up to ~600 K (Barman et al., 2001; Fortney et al., 2006; Goukenleuque et al., 2000; Iro et al., 2005; Seager and Sasselov, 2000; Sudarsky et al., 2003). Furthermore, as illustrated by **Figure 13**, the strong irradiation and expected synchronization of the planets' spin implies that strong inhomogeneities should exist in the atmosphere with in particular strong (~500 K) day–night and equator-to-pole differences in effective temperatures (Barman et al., 2005; Cooper and Showman, 2005; Iro et al., 2005; Showman and Guillot, 2002).

Figure 14 illustrates the expected structure for the atmosphere of HD 209458b from a modern, tridimensional global circulation model coupled with a one-dimensional radiative transfer algorithm (Parmentier et al., 2013). All the caveats concerning these overforced simulations discussed in **Section 10.16.2.4** of course also apply and add to the uncertainties stemming from the poorly known chemical composition. For example, the particular simulation of **Figure 14** assumes the presence of TiO in the atmosphere, which yields very high temperatures at low pressures on the dayside of the planet. It is not clear that this molecule is present or has condensed at deeper levels (see also Spiegel et al., 2009). Aside from that, the eastward equatorial circulation and the strong equator-to-pole gradient now appear to be a robust feature of these simulations (e.g., Parmentier et al., 2013; Rauscher and Menou, 2013; Showman and Guillot, 2002). As seen in **Figure 14**, the equatorial jet redistributes heat between the dayside and the nightside relatively efficiently at large pressures and on the equator, but this is not the case at the poles, and at low pressures, in line with the observational constraints (see **Section 10.16.2.10**).

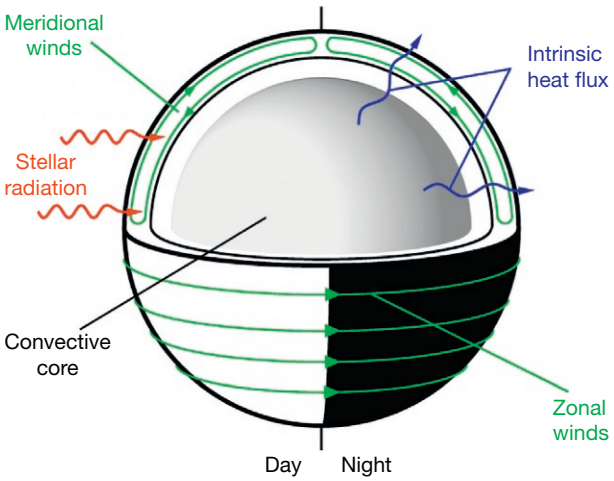


Figure 13 Conjectured dynamical structure of hot Jupiters (strongly irradiated extrasolar giant planets): At pressures larger than about 0.1 to 1 kbar, the intrinsic heat flux must be transported by convection. The convective core is at or near synchronous rotation with the star and has small latitudinal and longitudinal temperature variations. At lower pressures, a radiative envelope is present. The top part of the atmosphere is penetrated by the stellar light on the dayside. The spatial variation in insolation should drive winds that transport heat from the dayside to the nightside. Reproduced from Showman AP and Guillot T (2002) Atmospheric circulation and tides of "51 Pegasus b-like" planets. *Astronomy and Astrophysics* 385: 166–180.

These strong temperature variations must influence at some point the cooling and contraction histories of hot Jupiters. When opacities variations are not included, they result in more loss of the intrinsic heat and a faster contraction than when assuming that the stellar irradiation is homogeneously redistributed across the planetary surface (Budaj et al., 2012; Guillot, 2010; Guillot and Showman, 2002; Spiegel and Burrows, 2013). However, given other uncertainties (e.g., on the chemical composition and opacities to be used), this has been neglected in planetary evolution models thus far.

10.16.4.3.2 Thermal evolution and inferred compositions

We have seen in **Figure 8** that the measured masses and radii of transiting planets can be globally explained in the framework of an evolution model including the strong stellar irradiation and the presence of a variable mass of heavy elements, either in the form of a central core or spread in the planet interior. However, when analyzing the situation for each planet, it appears that several planets are too large to be reproduced by standard models, that is, models using the most up-to-date EOS, opacities, and atmospheric boundary conditions and assuming that the planetary luminosity governing its cooling is taken solely from the lost gravitational potential energy (see **Section 10.16.3.4**).

Figure 15 illustrates the situation for the particular case of HD 209458b: unless using an unrealistically hot atmosphere, or arbitrarily increasing the internal opacity, or decreasing the helium content, one cannot reproduce the observed radius that is 10–20% larger than calculated using standard models (Baraffe et al., 2003; Bodenheimer et al., 2001, 2003; Guillot and Showman, 2002). The fact that the measured radius corresponds to a low-pressure (~mbar) level while the calculated radius corresponds to a level near 1 bar is not negligible (Burrows et al., 2003) but too small to account for the difference. This is problematic because while it is easy to invoke the presence of a massive core to explain the small size of a planet, a large size such as that of HD 209458b requires an additional energy source or significant modifications in the data/physics involved.

The discovery of many transiting hot Jupiters has shown that this phenomenon is widespread, with at least a third of them being oversized compared to predictions from the standard evolution of a solar composition planet with no core (Guillot, 2008; Guillot et al., 2006; Laughlin et al., 2011). Numerous explanations have been put forward to explain this large size. The first ones, invoking tidal dissipation of eccentricity (Bodenheimer et al., 2001) or inclination (Winn and Holman, 2005), imply that orbital energy is transferred to the planet. These are generally too short-lived (e.g., Leconte et al., 2010) or of a low probability of occurrence (Levrard et al., 2007). The second ones posit that part of the irradiation energy is transferred into kinetic energy and is then dissipated deeper into the planet. This is the case of weather noise (Showman and Guillot, 2002), ohmic dissipation (Batygin

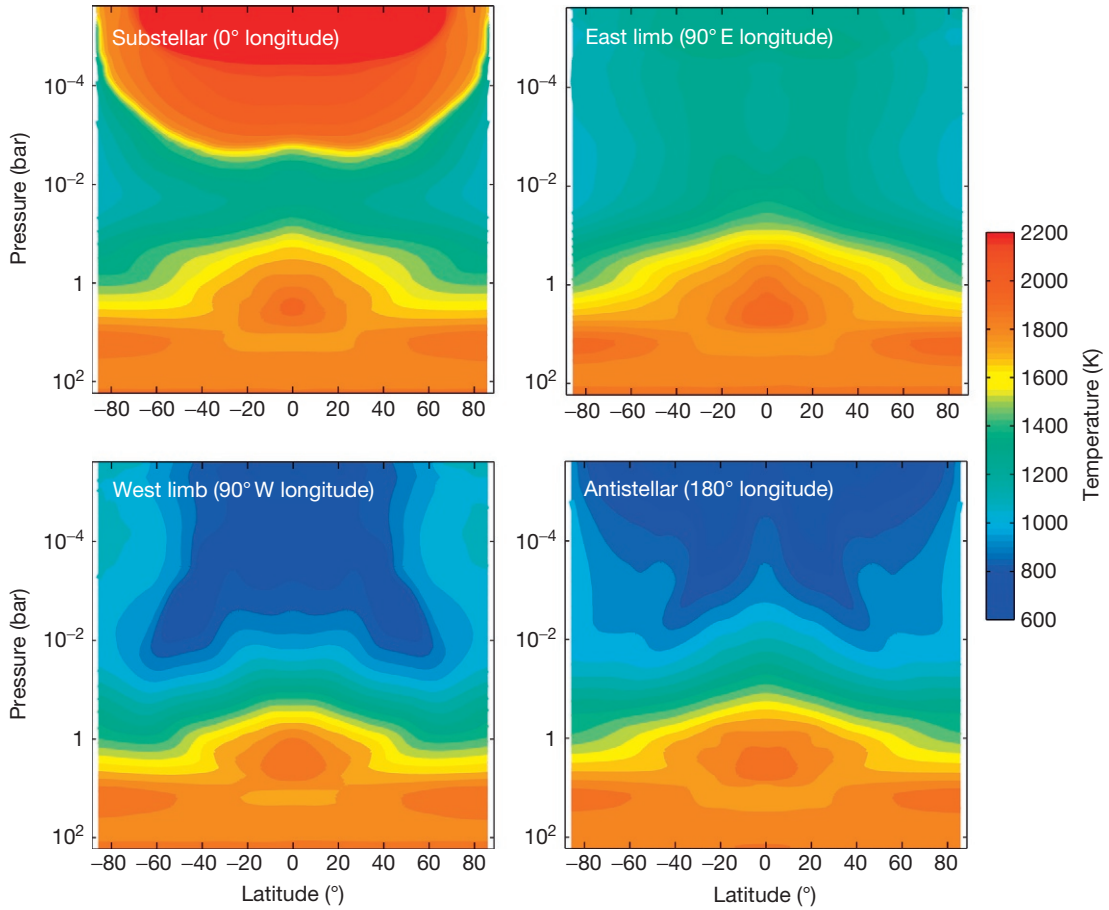


Figure 14 Simulated temperatures as a function of latitude (0° at the equator, ±90° at the poles) and pressure for HD 209458b obtained from a global 3D global circulation model. The four panels correspond to cuts from pole to pole at different longitudes (clockwise from the upper left): 0° (crossing the substellar point), 90° E (along the east limb), 180° E (crossing the antistellar point), and 270° E (along the west limb). A superrotating equatorial jet (0° latitude) is present and characterized by warmer temperatures than its surrounding at pressures of a few bars. The globally west to east circulation is responsible for a pronounced asymmetry between the warmer east limb and colder west limb. Reproduced from Parmentier V, Showman AP, and Lian Y (2013) 3D mixing in hot Jupiter's atmospheres. I. Application to the day/night cold trap in HD 209458b. *Astronomy and Astrophysics* 558: A91. Courtesy of V. Parmentier.

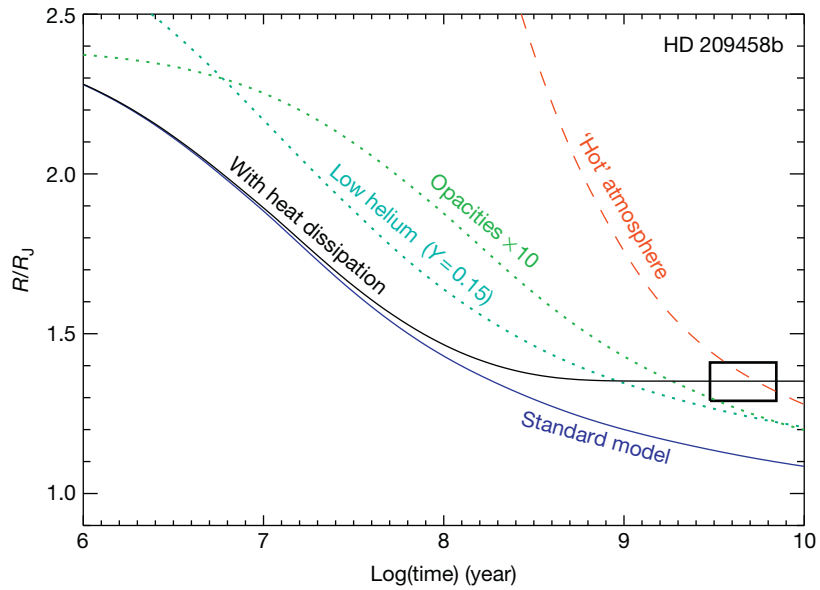


Figure 15 The contraction of HD 209458b as a function of time can be compared to its measured radius and inferred age shown by the black box. Standard models (blue curve) for the evolution of that 0.69 M_J planet generally yield a radius that is too small compared to the observations, even for a solar composition and no central core (a larger core and – in most cases – larger amounts of heavy elements in the planet imply an even smaller size for a given age). Unrealistically low helium abundances or high opacities models lead to evolution tracks that barely cross the observational box. A possibility is that heat is dissipated into the deep interior by stellar tides, either related to a nonzero orbital eccentricity forced by an unseen companion or because of a constant transfer of angular momentum from the heated atmosphere to the interior (black curve). Alternatively, the atmosphere may be hotter than predicted due to heating by strong zonal winds and shear instabilities (red curve).

and Stevenson, 2010), thermal tides (Arras and Socrates, 2010), and turbulent burial (Youdin and Mitchell, 2010) models. These mechanisms appear quite promising as they are long-lived and generally require only a small fraction of order 1% or less of the irradiation luminosity to be transported and dissipated at deeper levels to explain the observed planets (Guillot and Showman, 2002). Finally, a third class of models is based on a reduced cooling, either through an ad hoc increase of opacities (Burrows et al., 2007) or inefficient heat transport due to semiconvection (Chabrier and Baraffe, 2007). Validating these models is becoming possible, thanks to a large number of planets allowing statistical tests (see Laughlin et al., 2011), but will require further work.

In any case, the fact that a large number of planets are oversized lends weight to a mechanism that would apply to each planet. Masses of heavy elements can then be derived by imposing that all planets should be fitted by the same model with the same hypotheses. This can be done by inverting the results of **Figure 8**, as proposed by Guillot et al. (2006). On the basis of this hypothesis, **Figure 16** shows that some of the hot Jupiters contain a large fraction of heavy elements in their interior and that this fraction is correlated with the metallicity of the parent star. The large fraction of heavy elements is inferred both for relatively small giant planets (i.e., Saturn mass) and for planets that are several times the mass of Jupiter. A typical archetype of the first is HD 149026b that must contain around 70 $M_\oplus$ of heavy elements, a conclusion that is hard to escape because of the low total mass and high irradiation of the planet (see Fortney et al., 2006; Ikoma et al., 2006). For the latter, CoRoT-20b appears to push the models to the limit, with predicted masses of heavy elements in excess of 400 $M_\oplus$ (Deleuil et al., 2012).

The correlation between mass of heavy elements in hot Jupiters and stellar metallicity first obtained by Guillot et al. (2006) appears to stand the trial of time (Burrows et al., 2007; Guillot, 2008; Laughlin et al., 2011; Moutou et al., 2013) and remains valid when applied to planets with low irradiation levels that do not require additional physics to explain their large sizes (Miller and Fortney, 2011). It is important to realize that simply accreting slightly more metal-rich gas with the composition of the parent star would lead to a much smaller increase of a few percent at most. This correlation requires an efficient mechanism to collect solids in the protoplanetary disk and bring them into the hot Jupiters, something that is just beginning to be included into planet formation models (Mordasini et al., 2012).

Another intriguing possibility concerning hot Jupiters is that of a sustained mass loss due to the high irradiation dose that the planets receive. Indeed, this effect was predicted (Burrows and Lunine, 1995; Guillot et al., 1996; Lammer et al., 2003) and detected (Bourrier et al., 2013; Fossati et al., 2010; Vidal-Madjar et al., 2003). While its magnitude is still uncertain, it appears to have sculpted the population of planets in very close orbits around their star (e.g., Lopez et al., 2012).

The harvest of the Kepler and CoRoT missions opens the possibility to extend these studies to smaller planets. These objects are especially interesting but pose difficult problems in terms of structure because depending on their formation history, precise composition, and location, they may be fluid or solid or they may even possess a global liquid ocean (see Kuchner, 2003; Léger et al., 2004).

10.16.5 Implications for Planetary Formation Models

The giant planets in our solar system have in common a large mass of hydrogen and helium, but they are obviously quite different in their appearances, compositions, and internal

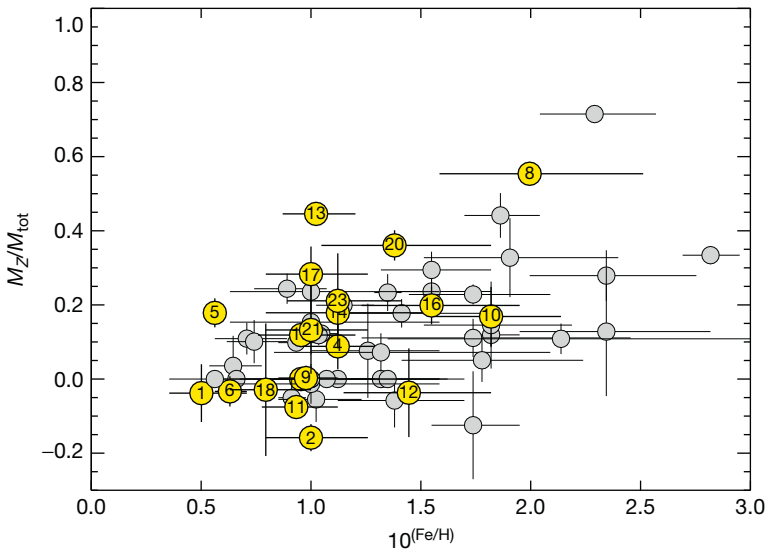


Figure 16 Mass fraction of heavy elements in the planets as a function of the metal content of their parent star expressed in solar units (e.g., $10^{[Fe/H]} = 3$ implies that the star contains three times more iron than our Sun). The evolution model assumes that 0.25% of the incoming irradiation flux is dissipated at the planet's center. Circles that are labeled 1–23 correspond to the CoRoT giant planets. Gray symbols correspond to a subset of known transiting systems (Guillot, 2008; Laughlin et al., 2011). Unphysical negative values for M_Z correspond to insufficient heat sources leading to a radius that is larger than observed. Reproduced from Moutou C, Deleuil M, Guillot T, et al. (2013) CoRoT: Harvest of the exoplanet program. *Icarus* 226: 1625–1634.

structures. Although studies cannot be conducted with the same level of details, we can safely conclude that extrasolar planets show a greater variety of compositions and structures, and imagine that their appearances differ even more significantly.

A parallel study of the structures of our giant planets and of the giant planets orbiting around other stars should provide us with key information regarding planet formation in the next decade or so. But, already, some conclusions, some of them robust, others still tentative, can be drawn:

The giant planets formed in circumstellar disks, before these were completely dissipated: This is a relatively obvious consequence of the fact that the giant planets are mostly made of hydrogen and helium: these elements had to be acquired when they were still present in the disk. Because the observed lifetime of gaseous circumstellar disks is of the order of a few million years, this implies that these planets formed (i.e., acquired most of their final masses) in a few million years also, quite faster than terrestrial planets in the solar system.

The giant planets migrated: The observed orbital distribution of extrasolar planets and the presence of planets extremely close to their star is generally taken as a strong evidence for an inward migration of planets, and various mechanisms have been proposed for that (see Alibert et al., 2005; Ida and Lin, 2004a; Moorhead and Adams, 2005, etc.). Separately, it was shown that several properties of our solar system can be explained if Jupiter, Saturn, Uranus, and Neptune ended the early formation phase in the presence of a disk with quasi-circular orbit, and with Saturn, Uranus, and Neptune significantly closer to the Sun than they are now, and that these three planets subsequently migrated outward (Tsiganis et al., 2005).

Accretion played a key role for the giant planet formation: Although formation by direct gas instability still remains a possibility (e.g., Boley et al., 2011; Helled and Bodenheimer, 2011), several indications point toward a formation of the giant planets that is dominated by accretion of heavy elements: First, Jupiter, Saturn, Uranus, and Neptune are all significantly enriched in heavy elements compared to the Sun. This feature can be reproduced by core-accretion models, for Jupiter and Saturn at least (Alibert et al., 2005). Second, the probability to find a giant planet around a solar-type star (with stellar type F, G, or K) is a strongly rising function of stellar metallicity (Fischer and Valenti, 2005; Gonzalez, 1998; Santos et al., 2004), a property that is also well-reproduced by standard core-accretion models (Alibert et al., 2005; Ida and Lin, 2004b). Third, the large masses of heavy elements inferred in some transiting extrasolar planets as well as the apparent correlation between mass of heavy elements in the planet and stellar metallicity (Burrows et al., 2007; Guillot, 2008; Guillot et al., 2006; Laughlin et al., 2011) are a strong indication that accretion was possible and that it was furthermore efficient.

The giant planets were enriched in heavy elements by core accretion, planetesimal delivery, and/or formation in an enriched protoplanetary disk: The giant planets in our solar system are unambiguously enriched in heavy elements compared to the Sun, both globally and when considering their atmosphere. This may also be the case of extrasolar planets, although the evidence is still tenuous. The accretion of a central core can explain part of the global enrichment, but not that of the atmosphere. The accretion of planetesimals may be a possible solution, but in the case of Jupiter at least, the rapid drop in accretion efficiency as the planet reaches appreciable masses ($\sim 100\ M_\oplus$ or so) implies that such an enrichment would have originally concerned only very deep layers and would require a relatively efficient upper mixing of these elements and possibly an erosion of the central core (Guillot et al., 2004; Wilson and Militzer, 2012).

Although not unambiguously explained, the fact that Jupiter is also enriched in noble gases compared to the Sun is a key observation to understand some of the processes occurring in the early solar system. Indeed, noble gases are trapped into solids only at very low temperatures, and this tells us that either most of the solids that formed Jupiter were formed at very low temperature to be able to trap gases such as argon, probably as clathrates (Gautier et al., 2001; Hersant et al., 2004; Mousis et al., 2012), or that the planet formed in an enriched disk as it was being evaporated (Guillot and Hueso, 2006). The fact that in Jupiter, argon, krypton, and xenon have a comparable enrichment over the solar value within the error bars (see Lodders, 2008, and **Table 3**) slightly favors the latter explanation.

10.16.6 Future Prospects

We have shown that the compositions and structures of the giant planets remain very uncertain. This is an important problem when attempting to understand and constrain the formation of planets, and the origins of the solar system. However, the parallel study of the giant planets in our solar system by space missions, such as Galileo and Cassini, and of extrasolar planets by both ground-based and space programs has led to rapid improvements in the field, with in particular a precise determination of the composition of Jupiter's troposphere, and constraints on the compositions of a dozen of extrasolar planets.

Improvements in our knowledge of the giant planets require a variety of efforts. Fortunately, nearly all of these are addressed at least partially by adequate projects in the next few years. The efforts that are necessary thus include (but are not limited to):

- Continue progresses on EOSs in order to obtain reliable results that can be used for a wide range of temperatures and pressures in the astrophysical context. This should be done with the help of laboratory experiments, for instance, by using powerful lasers such as the NIF in the United States and the MégaJoule laser in France. Extensive numerical calculations should be performed as well, in particular with mixtures of elements.
- Calculate phase diagrams for a variety of mixtures, in particular involving superionic water, rocks, and iron (and hydrogen). The hydrogen–helium phase diagram should

also be refined because it is critical to understand the evolution and structure of Jupiter and Saturn.

- Have a better yardstick to measure solar and protosolar compositions. This has not been fully addressed by the Genesis mission and may require another mission and/or progresses in modeling the Sun's composition.
- Improve the measurement of Jupiter's gravity and magnetic fields, and determine the abundance of water in the deep atmosphere. This will be done by the Juno mission (Bolton, 2010) that is to arrive at Jupiter in 2016 and whose polar orbit skimming a mere 5000 km above the cloud tops should enable exquisite measurements of these quantities.
- Measure with high-precision Saturn's gravity field. Saturn's gravitational moments have already been improved, but an important increase in accuracy can be obtained as part of the Cassini Solstice mission (Spilker, 2012), while the spacecraft plunges onto the planet. This should lead to better constraints and possibly a determination of whether the interior of Saturn rotates as a solid body.
- Pursue the discovery of transiting extrasolar planets including some with longer orbital periods and around bright stars. A large number of these objects will enable detailed statistical studies that will be key in understanding this population of objects.
- Develop consistent models for the formation, evolution, present structure, and magnetic field of Uranus and Neptune, in order to understand ice giants as a class of planets.
- It would be highly desirable to send a probe similar to the Galileo probe into Saturn's atmosphere (e.g., Marty et al., 2009). The comparison of the abundance of noble gases would discriminate between different models of the enrichment of the giant planets, and the additional measurement of key isotopic ratio would provide further tests to understand our origins.
- In the long term, a mission to the ice giants Uranus or Neptune would bring new views of these fascinating planets and help to complete our knowledge of the outer solar system.

Clearly, there is a lot of work on the road, but the prospects for a much improved knowledge of the giant planets and their formation are bright.

Acknowledgments

The manuscript improved significantly, thanks to the insightful reviews of Nadine Nettelmann and another reviewer. The authors also wish to thank Vivien Parmentier, Leigh Fletcher, Didier Saumon, Emmanuel Lellouch, Paul Loubeyre, Ravit Helled, Bill Hubbard, Erich Karkoschka, Imke De Pater, and Miguel Morales for very useful comments and suggestions.

References

Acuna MH, Connerney JEP, and Ness NF (1983) The Z3 zonal harmonic model of Saturn's magnetic field: Analyses and implications. *Journal of Geophysical Research* 88: 8771–8778.

Alibert Y, Mordasini C, Benz W, and Winisdoerffer C (2005) Models of giant planet formation with migration and disc evolution. *Astronomy and Astrophysics* 434: 343–353.

Anderson JD, Campbell JK, Jacobson RA, Sweetnam DN, and Taylor AH (1987) Radio science with Voyager 2 at Uranus – Results on masses and densities of the planet and five principal satellites. *Journal of Geophysical Research* 92: 14877–14883.

Anderson JD and Schubert G (2007) Saturn's gravitational field, internal rotation, and interior structure. *Science* 317: 1384–1387.

Arras P and Socrates A (2010) Thermal tides in fluid extrasolar planets. *Astrophysical Journal* 714: 1–12.

Atreya SK, Mahaffy PR, Niemann HB, Wong MH, and Owen TC (2003) Composition and origin of the atmosphere of Jupiter – An update and implications for the extrasolar giant planets. *Planetary and Space Science* 51: 105–112.

Bahcall JN, Pinsonneault MH, and Wasserburg GJ (1995) Solar models with helium and heavy-element diffusion. *Reviews of Modern Physics* 67: 781–808.

Baines KH and Smith HW (1990) The atmospheric structure and dynamical properties of Neptune derived from ground-based and IUE spectrophotometry. *Icarus* 85: 65–108.

Bakos GÁ, Howard AW, Noyes RW, et al. (2009) HAT-P-13b, c: A transiting hot Jupiter with a massive outer companion on an eccentric orbit. *Astrophysical Journal* 707: 446–456.

Baraffe I, Chabrier G, Barman TS, Allard F, and Hauschildt PH (2003) Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. *Astronomy and Astrophysics* 402: 701–712.

Baraffe I, Chabrier G, Barman TS, Selsis F, Allard F, and Hauschildt PH (2005) Hot-Jupiters and hot-Neptunes: A common origin? *Astronomy and Astrophysics* 436: L47–L51.

Barman TS, Hauschildt PH, and Allard F (2001) Irradiated planets. *Astrophysical Journal* 556: 885–895.

Barman TS, Hauschildt PH, and Allard F (2005) Phase-dependent properties of extrasolar planet atmospheres. *Astrophysical Journal* 632: 1132–1139.

Batygin K, Bodenheimer P, and Laughlin G (2009) Determination of the interior structure of transiting planets in multiple-planet systems. *Astrophysical Journal Letters* 704: L49–L53.

Batygin K and Stevenson DJ (2010) Inflating hot Jupiters with Ohmic dissipation. *Astrophysical Journal Letters* 714: L238–L243.

Bercovici D and Schubert G (1987) Jovian seismology. *Icarus* 69: 557–565.

Bethkenhagen M, French M, and Redmer R (2013) Equation of state and phase diagram of ammonia at high pressures from ab initio simulations. *Journal of Chemical Physics* 138(23): 234504.

Bodenheimer P, Laughlin G, and Lin DNC (2003) On the radii of extrasolar giant planets. *Astrophysical Journal* 592: 555–563.

Bodenheimer P, Lin DNC, and Mardling RA (2001) On the tidal inflation of short-period extrasolar planets. *Astrophysical Journal* 548: 466–472.

Boley AC, Helled R, and Payne MJ (2011) The heavy-element composition of disk instability planets can range from sub- to super-Nebular. *Astrophysical Journal* 735: 30.

Bolton SJ (2010) The Juno mission. In: Barbieri C, Chakrabarti S, Coradini M, and Lazzarin M (Eds.) *IAU Symposium*, vol. 269, pp. 92–100.

Bonev SA, Militzer B, and Galli G (2004) Ab initio simulations of dense liquid deuterium: Comparison with gas-gun shock-wave experiments. *Physical Review B* 69(1): 014101.

Boriskov GV, Bykov AI, Il'Kaev RI, et al. (2005) Shock compression of liquid deuterium up to 109 GPa. *Physical Review B* 71(9): 092104.

Borysow A, Jorgensen UG, and Zheng C (1997) Model atmospheres of cool, low-metallicity stars: The importance of collision-induced absorption. *Astronomy and Astrophysics* 324: 185–195.

Bouchy F, Deleuil M, Guillot T, et al. (2011) Transiting exoplanets from the CoRoT space mission XV CoRoT–15b: A brown-dwarf transiting companion. *Astronomy and Astrophysics* 525: A68.

Bourrier V, Lecavelier des Etangs A, Dupuy H, et al. (2013) Atmospheric escape from HD 189733b observed in H I Lyman-α: Detailed analysis of HST/STIS September 2011 observations. *Astronomy and Astrophysics* 551: A63.

Briggs FH and Sackett PD (1989) Radio observations of Saturn as a probe of its atmosphere and cloud structure. *Icarus* 80: 77–103.

Budaj J, Hubeny I, and Burrows A (2012) Day and night side core cooling of a strongly irradiated giant planet. *Astronomy and Astrophysics* 537: A115.

Burrows A (2013) Spectra as windows into exoplanet atmospheres. ArXiv e-prints.

Burrows A, Guillot T, Hubbard WB, et al. (2000) On the radii of close-in giant planets. *Astrophysical Journal Letters* 534: L97–L100.

Burrows A, Hubeny I, Budaj J, and Hubbard WB (2007) Possible solutions to the radius anomalies of transiting giant planets. *Astrophysical Journal* 661: 502–514.

Burrows A and Lunine J (1995) Extrasolar planets – Astronomical questions of origin and survival. *Nature* 378: 333.

Burrows A, Marley M, Hubbard WB, et al. (1997) A nongray theory of extrasolar giant planets and brown dwarfs. *Astrophysical Journal* 491: 856.

Burrows A, Sudarsky D, and Hubbard WB (2003) A theory for the radius of the transiting giant planet HD 209458b. *Astrophysical Journal* 594: 545–551.

Busse FH (1978) Magnetohydrodynamics of the Earth's dynamo. *Annual Review of Fluid Mechanics* 10: 435–462.

Caillabet L, Mazevet S, and Loubeyre P (2011) Multiphase equation of state of hydrogen from ab initio calculations in the range 0.2 to 5 g/cc up to 10 eV. *Physical Review B* 83(9): 094101.

Campbell JK and Synnott SP (1985) Gravity field of the Jovian system from Pioneer and Voyager tracking data. *Astronomical Journal* 90: 364–372.

Cavazzoni C, Chiarotti GL, Scandolo S, Tosatti E, Bernasconi M, and Parrinello M (1999) Superionic and metallic states of water and ammonia at giant planet conditions. *Science* 283: 44.

Cecconi B and Zarka P (2005) Model of a variable radio period for Saturn. *Journal of Geophysical Research (Space Physics)* 110: 12203.

Chabrier G and Ashcroft NW (1990) Linear mixing rule in screened binary ionic mixtures. *Physical Review A* 42: 2284–2291.

Chabrier G and Baraffe I (2007) Heat transport in giant (exo)planets: A new perspective. *Astrophysical Journal Letters* 661: L81–L84.

Chabrier G, Saumon D, and Winisdoerffer C (2007) Hydrogen and helium at high density and astrophysical implications. *Astrophysics and Space Science* 307: 263–267.

Chandrasekhar S (1939) *An Introduction to the Study of Stellar Structure*. Chicago, IL: University of Chicago Press.

Charbonneau D, Brown TM, Noyes RW, and Gilliland RL (2002) Detection of an extrasolar planet atmosphere. *Astrophysical Journal* 568: 377–384.

Christensen-Dalsgaard J, Dappen W, Ajukov SV, et al. (1996) The current state of solar modeling. *Science* 272: 1286–1292.

Cohen ER and Taylor BN (1987) The 1986 adjustment of the fundamental physical constants. *Reviews of Modern Physics* 59: 1121–1148.

Collins GW, da Silva LB, Celliers P, et al. (1998) Measurements of the equation of state of deuterium at the fluid insulator-metal transition. *Science* 281: 1178–1181.

Connerney JEP, Ness NF, and Acuna MH (1982) Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* 298: 44–46.

Conrath BJ and Gautier D (2000) Saturn helium abundance: A reanalysis of Voyager measurements. *Icarus* 144: 124–134.

Conrath BJ, Gautier D, Lindal GF, Samuelson RE, and Shaffer WA (1991) The helium abundance of Neptune from Voyager measurements. *Journal of Geophysical Research* 96: 18907.

Conrath B, Hanel R, Gautier D, Marten A, and Lindal G (1987) The helium abundance of Uranus from Voyager measurements. *Journal of Geophysical Research* 92: 15003–15010.

Cooper CS and Showman AP (2005) Dynamic meteorology at the photosphere of HD 209458b. *Astrophysical Journal Letters* 629: L45–L48.

Crossfield IJM, Barman T, Hansen BMS, Tanaka I, and Kodama T (2012) Re-evaluating WASP-12b: Strong emission at 2.315 μm, deeper occultations, and an isothermal atmosphere. *Astrophysical Journal* 760: 140.

Crouzet N, McCullough PR, Burke C, and Long D (2012) Transmission spectroscopy of exoplanet XO-2b observed with Hubble Space Telescope NICMOS. *Astrophysical Journal* 761: 7.

da Silva LB, Celliers P, Collins GW, et al. (1997) Absolute equation of state measurements on shocked liquid deuterium up to 200 GPa (2 Mbar). *Physical Review Letters* 78: 483–486.

Davies ME, Abalakin VK, Bursa M, Lederle T, and Lieske JH (1986) Report of the IAU/IAG/COSPAR working group on cartographic coordinates and rotational elements of the planets and satellites – 1985. *Celestial Mechanics* 39: 103–113.

de Pater I, Romani PN, and Atreya SK (1991) Possible microwave absorption by H2S gas in Uranus' and Neptune's atmospheres. *Icarus* 91: 220–233.

de Pater I, Sromovsky LA, Hammel HB, et al. (2011) Post-equinox observations of Uranus: Berg's evolution, vertical structure, and track towards the equator. *Icarus* 215: 332–345.

Deleuil M, Bonomo AS, Ferraz-Mello S, et al. (2012) Transiting exoplanets from the CoRoT space mission. XX. CoRoT-20b: A very high density, high eccentricity transiting giant planet. *Astronomy and Astrophysics* 538: A145.

Deming D, Wilkins A, McCullough P, et al. (2013) Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophysical Journal* 774: 95.

Désert J-M, Lecavelier des Etangs A, Hébrard G, et al. (2009) Search for carbon monoxide in the atmosphere of the transiting exoplanet HD 189733b. *Astrophysical Journal* 699: 478–485.

Desjarlais MP (2003) Density-functional calculations of the liquid deuterium Hugoniot, reshock, and reverberation timing. *Physical Review B* 68(6): 064204.

Fegley BJ and Lodders K (1994) Chemical models of the deep atmospheres of Jupiter and Saturn. *Icarus* 110: 117–154.

Feuchtgruber H, Lellouch E, Orton G, et al. (2013) The D/H ratio in the atmospheres of Uranus and Neptune from Herschel-PACS observations. *Astronomy and Astrophysics* 551: A126.

Fischer DA and Valenti J (2005) The planet-metallicity correlation. *Astrophysical Journal* 622: 1102–1117.

Fletcher LN, Baines KH, Momary TW, et al. (2011) Saturn's tropospheric composition and clouds from Cassini/VIMS 4.6-5.1 μm nightside spectroscopy. *Icarus* 214: 510–533.

Fletcher LN, Orton GS, Teanby NA, and Irwin PGJ (2009a) Phosphine on Jupiter and Saturn from Cassini/CIRS. *Icarus* 202: 543–564.

Fletcher LN, Orton GS, Teanby NA, Irwin PGJ, and Bjoraker GL (2009b) Methane and its isotopologues on Saturn from Cassini/CIRS observations. *Icarus* 199: 351–367.

Fortney JJ and Hubbard WB (2003) Phase separation in giant planets: Inhomogeneous evolution of Saturn. *Icarus* 164: 228–243.

Fortney JJ, Ikoma M, Nettelmann N, Guillot T, and Marley MS (2011) Self-consistent model atmospheres and the cooling of the solar system's giant planets. *Astrophysical Journal* 729: 32.

Fortney JJ and Nettelmann N (2010) The interior structure, composition, and evolution of giant planets. *Space Science Reviews* 152: 423–447.

Fortney JJ, Saumon D, Marley MS, Lodders K, and Freedman RS (2006) Atmosphere, interior, and evolution of the metal-rich transiting planet HD 149026b. *Astrophysical Journal* 642: 495–504.

Fortov VE, Ilkaev RI, Arinin VA, et al. (2007) Phase transition in a strongly nonideal deuterium plasma generated by quasi-isentropical compression at megabar pressures. *Physical Review Letters* 99(18): 185001.

Fossati L, Haswell CA, Froning CS, et al. (2010) Metals in the exosphere of the highly irradiated planet WASP-12b. *Astrophysical Journal Letters* 714: L222–L227.

Freedman RS, Marley MS, and Lodders K (2008) Line and mean opacities for ultracool dwarfs and extrasolar planets. *Astrophysical Journal* 174: 504–513.

French M, Becker A, Lorenzen W, et al. (2012) Ab initio simulations for material properties along the Jupiter adiabat. *Astrophysical Journal* 202(supplement): 5.

French M, Mattsson TR, Nettelmann N, and Redmer R (2009) Equation of state and phase diagram of water at ultrahigh pressures as in planetary interiors. *Physical Review B* 79(5): 054107.

Gastine T, Wicht J, and Aurnou JM (2013) Zonal flow regimes in rotating anelastic spherical shells: An application to giant planets. *Icarus* 225: 156–172.

Gaulme P, Schmider F-X, Gay J, Guillot T, and Jacob C (2011) Detection of Jovian seismic waves: A new probe of its interior structure. *Astronomy and Astrophysics* 531: A104.

Gautier D, Conrath BJ, Owen T, De Pater I, and Atreya SK (1995) The troposphere of Neptune. In: *Neptune and Triton*, pp. 547–611. Tucson, AZ: University of Arizona Press.

Gautier D, Hersant F, Mousis O, and Lunine JI (2001) Erratum: Enrichments in volatiles in Jupiter: A new interpretation of the Galileo measurements. *Astrophysical Journal Letters* 559: L183.

Giampieri G, Dougherty MK, Smith EJ, and Russell CT (2006) A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* 441: 62–64.

Gibson NP, Pont F, and Aigrain S (2011) A new look at NICMOS transmission spectroscopy of HD 189733, GJ-436 and XO-1: No conclusive evidence for molecular features. *Monthly Notices of the Royal Astronomical Society* 411: 2199–2213.

Gierasch PJ, Ingersoll AP, Banfield D, et al. (2000) Observation of moist convection in Jupiter's atmosphere. *Nature* 403: 628–630.

Glatzmaier G, Evonuk M, and Rogers T (2009) Differential rotation in giant planets maintained by density-stratified turbulent convection. *Geophysical and Astrophysical Fluid Dynamics* 103: 31–51.

Gonzalez G (1998) Spectroscopic analyses of the parent stars of extrasolar planetary system candidates. *Astronomy and Astrophysics* 334: 221–238.

Goukenleuque C, Bézard B, Joguet B, Lellouch E, and Freedman R (2000) A radiative equilibrium model of 51 Peg b. *Icarus* 143: 308–323.

Guervilly C, Cardin P, and Schaeffer N (2012) A dynamo driven by zonal jets at the upper surface: Applications to giant planets. *Icarus* 218: 100–114.

Guillot T (1999a) A comparison of the interiors of Jupiter and Saturn. *Planetary and Space Science* 47: 1183–1200.

Guillot T (1999b) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77.

Guillot T (2005) The interiors of giant planets: Models and outstanding questions. *Annual Review of Earth and Planetary Sciences* 33: 493–530.

Guillot T (2008) The composition of transiting giant extrasolar planets. *Physica Scripta, Volume T* 130(1): 014023.

Guillot T (2010) On the radiative equilibrium of irradiated planetary atmospheres. *Astronomy and Astrophysics* 520: A27.

Guillot T, Burrows A, Hubbard WB, Lunine JI, and Saumon D (1996) Giant planets at small orbital distances. *Astrophysical Journal Letters* 459: L35–L39.
Guillot T and Hueso R (2006) The composition of Jupiter: Sign of a (relatively) late formation in a chemically evolved protosolar disc. *Monthly Notices of the Royal Astronomical Society* 367: L47–L51.
Guillot T, Santos NC, Pont F, Iro N, Melo C, and Ribas I (2006) A correlation between the heavy element content of transiting extrasolar planets and the metallicity of their parent stars. *Astronomy and Astrophysics* 453: L21–L24.
Guillot T and Showman AP (2002) Evolution of "51 pegasus b-like" planets. *Astronomy and Astrophysics* 385: 156–165.
Guillot T, Stevenson DJ, Hubbard WB, and Saumon D (2004) The interior of Jupiter. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites, and Magnetosphere*, vol. 1, pp. 35–57. Cambridge: Cambridge University Press.
Gulkis S, Janssen MA, and Olsen ET (1978) Evidence for the depletion of ammonia in the Uranus atmosphere. *Icarus* 34: 10–19.
Gurnett DA, Kurth WS, Hospodarsky GB, et al. (2005) Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* 307: 1255–1259.
Hammel HB, de Pater I, Gibbard SG, Lockwood GW, and Rages K (2005) New cloud activity on Uranus in 2004: First detection of a southern feature at 2.2 μm. *Icarus* 175: 284–288.
Hansen BMS (2008) On the absorption and redistribution of energy in irradiated planets. *Astrophysical Journal* 179(supplement): 484–508.
Harrington J, Hansen BM, Luszcz SH, et al. (2006) The phase-dependent infrared brightness of the extrasolar planet υ Andromedae b. *Science* 314: 623–626.
Hedman MM and Nicholson PD (2013) Kronoseismology: Using density waves in Saturn's C ring to probe the planet's interior. *Astronomical Journal* 146: 12.
Heimpel M and Gómez Pérez N (2011) On the relationship between zonal jets and dynamo action in giant planets. *Geophysical Research Letters* 38: 14201.
Helled R (2011) Constraining Saturn's core properties by a measurement of its moment of inertia – Implications to the Cassini Solstice Mission. *Astrophysical Journal Letters* 735: L16.
Helled R, Anderson JD, Podolak M, and Schubert G (2011a) Interior models of Uranus and Neptune. *Astrophysical Journal* 726: 15.
Helled R, Anderson JD, and Schubert G (2010) Uranus and Neptune: Shape and rotation. *Icarus* 210: 446–454.
Helled R, Anderson JD, Schubert G, and Stevenson DJ (2011b) Jupiter's moment of inertia: A possible determination by Juno. *Icarus* 216: 440–448.
Helled R and Bodenheimer P (2011) The effects of metallicity and grain growth and settling on the early evolution of gaseous protoplanets. *Icarus* 211: 939–947.
Helled R and Guillot T (2013) Interior models of Saturn: Including the uncertainties in shape and rotation. *Astrophysical Journal* 767: 113.
Hersant F, Gautier D, and Lunine JI (2004) Enrichment in volatiles in the giant planets of the Solar System. *Planetary and Space Science* 52: 623–641.
Hicks DG, Boehly TR, Celliers PM, et al. (2009) Laser-driven single shock compression of fluid deuterium from 45 to 220 GPa. *Physical Review B* 79(1): 014112.
Hubbard WB (1968) Thermal structure of Jupiter. *Astrophysical Journal* 152: 745–754.
Hubbard WB (1977) The Jovian surface condition and cooling rate. *Icarus* 30: 305–310.
Hubbard WB (1989) Structure and composition of giant planet interiors. In: *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 539–563. Tucson, AZ: University of Arizona Press.
Hubbard WB (1999) NOTE: Gravitational signature of Jupiter's deep zonal flows. *Icarus* 137: 357–359.
Hubbard WB (2013) Concentric Maclaurin spheroid models of rotating liquid planets. *Astrophysical Journal* 768: 43.
Hubbard WB, Guillot T, Marley MS, Burrows A, Lunine JI, and Saumon DS (1999) Comparative evolution of Jupiter and Saturn. *Planetary and Space Science* 47: 1175–1182.
Hubbard WB, Pearl JC, Podolak M, and Stevenson DJ (1995) The interior of Neptune. In: *Neptune and Triton*, pp. 109–138. Tucson, AZ: University of Arizona Press.
Hueso R, Sánchez-Lavega A, and Guillot T (2002) A model for large-scale convective storms in Jupiter. *Journal of Geophysical Research (Planets)* 107: 1–5.
Ida S and Lin DNC (2004a) Toward a deterministic model of planetary formation I. A desert in the mass and semimajor axis distributions of extrasolar planets. *Astrophysical Journal* 604: 388–413.
Ida S and Lin DNC (2004b) Toward a deterministic model of planetary formation. II. The formation and retention of gas giant planets around stars with a range of metallicities. *Astrophysical Journal* 616: 567–572.
Ikoma M, Guillot T, Genda H, Tanigawa T, and Ida S (2006) On the origin of HD 149026b. *Astrophysical Journal* 650: 1150–1159.
Ingersoll AP, Barnet CD, Beebe RF, et al. (1995) Dynamic meteorology of Neptune. In: *Neptune and Triton*, pp. 613–682. Tucson, AZ: University of Arizona Press.
Iorio L (2010) Juno, the angular momentum of Jupiter and the Lense-Thirring effect. *New Astronomy* 15: 554–560.
Iro N, Bézard B, and Guillot T (2005) A time-dependent radiative model of HD 209458b. *Astronomy and Astrophysics* 436: 719–727.
Irwin PGJ, Lellouch E, de Bergh C, et al. (2014) Line-by-line analysis of Neptune's near-IR spectrum observed with Gemini/NIFS and VLT/CRIRES. *Icarus* 227: 37–48.
Jackiewicz J, Nettelmann N, Marley M, and Fortney J (2012) Forward and inverse modeling for jovian seismology. *Icarus* 220: 844–854.
Jacobson RA (2009) The orbits of the Neptunian satellites and the orientation of the pole of Neptune. *Astronomical Journal* 137: 4322–4329.
Jacobson RA, Antreasian PG, Bordi JJ, et al. (2006) The gravity field of the Saturnian system from satellite observations and spacecraft tracking data. *Astrophysical Journal* 132: 2520–2526.
Karkoschka E and Tomasko MG (2011) The haze and methane distributions on Neptune from HST-STIS spectroscopy. *Icarus* 211: 780–797.
Kaspi Y, Showman AP, Hubbard WB, Aharonson O, and Helled R (2013) Atmospheric confinement of jet streams on Uranus and Neptune. *Nature* 497: 344–347.
Kippenhahn R and Weigert A (1994) *Stellar Structure and Evolution*, vol. XVI. Berlin: Springer-Verlag, 468 pp. 192 figs; Also Astronomy and Astrophysics Library.
Knudson MD and Desjarlais MP (2009) Shock compression of quartz to 1.6 TPa: Redefining a pressure standard. *Physical Review Letters* 103(22): 225501.
Knudson MD, Desjarlais MP, Lemke RW, et al. (2012) Probing the interiors of the Ice giants: Shock compression of water to 700 GPa and 3.8 g/cm^3. *Physical Review Letters* 108(9): 091102.
Knudson MD, Hanson DL, Bailey JE, Hall CA, Asay JR, and Deeney C (2004) Principal Hugoniot, reverberating wave, and mechanical reshock measurements of liquid deuterium to 400 GPa using plate impact techniques. *Physical Review B* 69(14): 144209.
Knutson HA, Charbonneau D, Allen LE, et al. (2007) A map of the day-night contrast of the extrasolar planet HD 189733b. *Nature* 447: 183–186.
Kramm U, Nettelmann N, Fortney JJ, Neuhäuser R, and Redmer R (2012) Constraining the interior of extrasolar giant planets with the tidal Love number k_2 using the example of HAT-P-13b. *Astronomy and Astrophysics* 538: A146.
Kuchner MJ (2003) Volatile-rich Earth-mass planets in the habitable zone. *Astrophysical Journal Letters* 596: L105–L108.
Kunde V, Hanel R, Maguire W, et al. (1982) The tropospheric gas composition of Jupiter's north equatorial belt /NH3, PH3, CH3D, GeH4, H2O/ and the Jovian D/H isotopic ratio. *Astrophysical Journal* 263: 443–467.
Lammer H, Selsis F, Ribas I, Guinan EF, Bauer SJ, and Weiss WW (2003) Atmospheric loss of exoplanets resulting from stellar X-ray and extreme-ultraviolet heating. *Astrophysical Journal Letters* 598: L121–L124.
Laughlin G, Crismani M, and Adams FC (2011) On the anomalous radii of the transiting extrasolar planets. *Astrophysical Journal Letters* 729: L7.
Leconte J and Chabrier G (2012) A new vision of giant planet interiors: Impact of double diffusive convection. *Astronomy and Astrophysics* 540: A20.
Leconte J, Chabrier G, Baraffe I, and Levrard B (2010) Is tidal heating sufficient to explain bloated exoplanets? Consistent calculations accounting for finite initial eccentricity. *Astronomy and Astrophysics* 516: A64.
Léger A, Selsis F, Sotin C, et al. (2004) A new family of planets? "Ocean-Planets" *Icarus* 169: 499–504.
Lellouch E, Bézard B, Fouchet T, Feuchtgruber H, Encrenaz T, and de Graauw T (2001) The deuterium abundance in Jupiter and Saturn from ISO-SWS observations. *Astronomy and Astrophysics* 370: 610–622.
Levrard B, Correia ACM, Chabrier G, Baraffe I, Selsis F, and Laskar J (2007) Tidal dissipation within hot Jupiters: A new appraisal. *Astronomy and Astrophysics* 462: L5–L8.
Lian Y and Showman AP (2010) Generation of equatorial jets by large-scale latent heating on the giant planets. *Icarus* 207: 373–393.
Lindal GF (1992) The atmosphere of Neptune – An analysis of radio occultation data acquired with Voyager 2. *Astronomical Journal* 103: 967–982.
Lindal GF, Sweetnam DN, and Eshleman VR (1985) The atmosphere of Saturn – An analysis of the Voyager radio occultation measurements. *Astronomical Journal* 90: 1136–1146.
Lindal GF, Wood GE, Levy GS, et al. (1981) The atmosphere of Jupiter – An analysis of the Voyager radio occultation measurements. *Journal of Geophysical Research* 86: 8721–8727.
Little B, Anger CD, Ingersoll AP, et al. (1999) Galileo images of lightning on Jupiter. *Icarus* 142: 306–323.
Liu J and Schneider T (2011) Convective generation of equatorial superrotation in planetary atmospheres. *Journal of Atmospheric Sciences* 68: 2742–2756.
Liu J, Schneider T, and Kaspi Y (2013) Predictions of thermal and gravitational signals of Jupiter's deep zonal winds. *Icarus* 224: 114–125.
Lodders K (2008) The solar argon abundance. *Astrophysical Journal* 674: 607–611.

Lodders K and Fegley B Jr. (1994) The origin of carbon monoxide in Neptune's atmosphere. *Icarus* 112: 368–375.

Lodders K, Palme H, and Gail H-P (2009) Abundances of the elements in the solar system. In: *Landolt Börnstein – Group VI Astronomy and Astrophysics Numerical Data and Functional Relationships in Science and Technology* Volume, 44.

Lopez ED, Fortney JJ, and Miller N (2012) How thermal evolution and mass-loss sculpt populations of super-Earths and sub-Neptunes: Application to the Kepler-11 system and beyond. *Astrophysical Journal* 761: 59.

Lorenzen W, Holst B, and Redmer R (2011) Metallization in hydrogen-helium mixtures. *Physical Review B* 84(23): 235109.

Loubeyre P, Brygoo S, Eggert J, et al. (2012) Extended data set for the equation of state of warm dense hydrogen isotopes. *Physical Review B* 86(14): 144115.

Loubeyre P, Letoullec R, and Pinceaux JP (1991) A new determination of the binary phase diagram of H_2-He mixtures at 296 K. *Journal of Physics: Condensed Matter* 3: 3183–3192.

Mardling RA (2010) The determination of planetary structure in tidally relaxed inclined systems. *Monthly Notices of the Royal Astronomical Society* 407: 1048–1069.

Marley MS, Fortney JJ, Hubickyj O, Bodenheimer P, and Lissauer JJ (2007) On the luminosity of young Jupiters. Astrophysical Journal 655, 541.

Marley MS, Gómez P, and Podolak M (1995) Monte Carlo interior models for Uranus and Neptune. *Journal of Geophysical Research* 100: 23349–23354.

Marley MS and Porco CC (1993) Planetary acoustic mode seismology – Saturn's rings. *Icarus* 106: 508.

Marley MS, Saumon D, Guillot T, et al. (1996) Atmospheric, evolutionary, and spectral models of the brown dwarf Gliese 229 B. *Science* 272: 1919–1921.

Marty B, Chaussidon M, Wiens RC, Jurewicz AJG, and Burnett DS (2011) A ^{15}N-poor isotopic composition for the solar system as shown by Genesis solar wind samples. *Science* 332: 1533.

Marty B, Guillot T, Coustenis A, et al. (2009) Kronos: Exploring the depths of Saturn with probes and remote sensing through an international mission. *Experimental Astronomy* 23: 947–976.

Mayor M and Queloz D (1995) A Jupiter-mass companion to a solar-type star. *Nature* 378: 355.

McMahon JM, Morales MA, Pierleoni C, and Ceperley DM (2012) The properties of hydrogen and helium under extreme conditions. *Reviews of Modern Physics* 84: 1607–1653.

Militzer B, Ceperley DM, Kress JD, Johnson JD, Collins LA, and Mazevet S (2001) Calculation of a deuterium double shock Hugoniot from ab initio simulations. *Physical Review Letters* 87(26), A265502.

Militzer B and Hubbard WB (2013) Ab initio equation of state for hydrogen-helium mixtures with recalibration of the giant-planet mass-radius relation. *Astrophysical Journal* 774: 148.

Militzer B, Hubbard WB, Vorberger J, Tamblyn I, and Bonev SA (2008) A massive core in Jupiter predicted from first-principles simulations. *Astrophysical Journal Letters* 688: L45–L48.

Miller N and Fortney JJ (2011) The heavy-element masses of extrasolar giant planets, revealed. *Astrophysical Journal Letters* 736: L29.

Mochalov MA, Il'kaev RI, Fortov VE, et al. (2012) Measurement of quasi-isentropic compressibility of helium and deuterium at pressures of 1500–2000 GPa. *Soviet Journal of Experimental and Theoretical Physics* 115: 614–625.

Moorhead AV and Adams FC (2005) Giant planet migration through the action of disk torques and planet-planet scattering. *Icarus* 178: 517–539.

Morales MA, Hamel S, Caspersen K, and Schwegler E (2013a) Hydrogen-helium demixing from first principles: From diamond anvil cells to planetary interiors. *Physical Review B* 87(17): 174105.

Morales MA, McMahon JM, Pierleoni C, and Ceperley DM (2013b) Nuclear quantum effects and nonlocal exchange-correlation functionals applied to liquid hydrogen at high pressure. *Physical Review Letters* 110(6): 065702.

Morales MA, Pierleoni C, and Ceperley DM (2010) Equation of state of metallic hydrogen from coupled electron-ion Monte Carlo simulations. *Physical Review E* 81(2): 021202.

Morales MA, Schwegler E, Ceperley D, Pierleoni C, Hamel S, and Caspersen K (2009) Phase separation in hydrogen-helium mixtures at Mbar pressures. *Proceedings of the National Academy of Sciences* 106: 1324.

Mordasini C (2013) Luminosity of young Jupiters revisited: Massive cores make hot planets. *Astronomy and Astrophysics* 558: A113.

Mordasini C, Alibert Y, Georgy C, Dittkrist K-M, Klahr H, and Henning T (2012) Characterization of exoplanets from their formation. II. The planetary mass-radius relationship. *Astronomy and Astrophysics* 547: A112.

Mousis O, Lunine JI, Madhusudhan N, and Johnson TV (2012) Nebular water depletion as the cause of Jupiter's low oxygen abundance. *Astrophysical Journal Letters* 751: L7.

Moutou C, Deleuil M, Guillot T, et al. (2013) CoRoT: Harvest of the exoplanet program. *Icarus* 226: 1625–1634.

Ness NF, Acuna MH, Behannon KW, Burlaga LF, Connerney JEP, and Lepping RP (1986) Magnetic fields at Uranus. *Science* 233: 85–89.

Ness NF, Acuna MH, Burlaga LF, Connerney JEP, and Lepping RP (1989) Magnetic fields at Neptune. *Science* 246: 1473–1478.

Nettelmann N, Becker A, Holst B, and Redmer R (2012) Jupiter models with improved ab initio hydrogen equation of state (H-REOS.2). *Astrophysical Journal* 750: 52.

Nettelmann N, Helled R, Fortney JJ, and Redmer R (2013a) New indication for a dichotomy in the interior structure of Uranus and Neptune from the application of modified shape and rotation data. *Planetary and Space Science* 77: 143–151.

Nettelmann N, Püstow R, and Redmer R (2013b) Saturn layered structure and homogeneous evolution models with different EOSs. *Icarus* 225: 548–557.

Noll KS and Larson HP (1991) The spectrum of Saturn from 1990 to 2230/cm – Abundances of AsH_3, CH_3D, CO, GeH_4, NH_3, and PH_3. *Icarus* 89: 168–189.

Noll KS, Larson HP, and Geballe TR (1990) The abundance of AsH_3 in Jupiter. *Icarus* 83: 494–499.

Owen T, Mahaffy P, Niemann HB, et al. (1999) A low-temperature origin for the planetesimals that formed Jupiter. *Nature* 402: 269–270.

Parmentier V and Guillot T (2014) A non-grey analytical model for irradiated atmospheres. I: Derivation. *Astronomy and Astrophysics* vol. 562, A133.

Parmentier V, Showman AP, and Lian Y (2013) 3D mixing in hot Jupiters atmospheres. I. Application to the day/night cold trap in HD 209458b. *Astronomy and Astrophysics* 558: A91.

Pearl JC and Conrath BJ (1991) The albedo, effective temperature, and energy balance of Neptune, as determined from Voyager data. *Journal of Geophysical Research* 96: 18921.

Podolak M, Hubbard WB, and Stevenson DJ (1991) Model of Uranus' interior and magnetic field. In: *Uranus*, pp. 29–61. Tucson, AZ: University of Arizona Press.

Podolak M, Podolak JI, and Marley MS (2000) Further investigations of random models of Uranus and Neptune. *Planetary and Space Science* 48: 143–151.

Podolak M, Weizman A, and Marley M (1995) Comparative models of Uranus and Neptune. *Planetary and Space Science* 43: 1517–1522.

Pont F, Sing DK, Gibson NP, Aigrain S, Henry G, and Husnoo N (2013) The prevalence of dust on the exoplanet HD 189733b from Hubble and Spitzer observations. *Monthly Notices of the Royal Astronomical Society* 432: 2917–2944.

Rages K, Hammel HB, and Lockwood GW (2002) A prominent apparition of Neptune's south polar feature. *Icarus* 159: 262–265.

Rauscher E and Menou K (2013) Three-dimensional atmospheric circulation models of HD 189733b and HD 209458b with consistent magnetic drag and ohmic dissipation. *Astrophysical Journal* 764: 103.

Read PL, Dowling TE, and Schubert G (2009) Saturn's rotation period from its atmospheric planetary-wave configuration. *Nature* 460: 608–610.

Redmer R, Mattsson TR, Nettelmann N, and French M (2011) The phase diagram of water and the magnetic fields of Uranus and Neptune. *Icarus* 211: 798–803.

Ribas I (2006) Masses and radii of low-mass stars: Theory versus observations. *Astrophysics and Space Science* 304: 89–92.

Robinson TD and Catling DC (2014) Common 0.1 bar tropopause in thick atmospheres set by pressure-dependent infrared transparency. *Nature Geoscience* 7: 12–15.

Rosenblum E, Garaud P, Traxler A, and Stellmach S (2011) Turbulent mixing and layer formation in double-diffusive convection: Three-dimensional numerical simulations and theory. *Astrophysical Journal* 731: 66.

Rossow WB (1978) Cloud microphysics – Analysis of the clouds of Earth, Venus, Mars, and Jupiter. *Icarus* 36: 1–50.

Roulston MS and Stevenson DJ (1995) Prediction of neon depletion in Jupiter's atmosphere. *EOS* 76: 343.

Russell CT and Dougherty MK (2010) Magnetic fields of the outer planets. *Space Science Reviews* 152: 251–269.

Salpeter EE (1973) On convection and gravitational layering in Jupiter and in stars of low mass. *Astrophysical Journal Letters* 181: L83.

Sanchez-Lavega A, Lecacheux J, Gomez JM, et al. (1996) Large-scale storms in Saturn's atmosphere during 1994. *Science* 271: 631–634.

Sano T, Ozaki N, Sakaiya T, et al. (2011) Laser-shock compression and Hugoniot measurements of liquid hydrogen to 55 GPa. *Physical Review B* 83(5): 054117.

Santos NC, Israelian G, and Mayor M (2004) Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation. *Astronomy and Astrophysics* 415: 1153–1166.

Sato B, Fischer DA, Henry GW, et al. (2005) The N2K consortium. II. A transiting hot Saturn around HD 149026 with a large dense core. *Astrophysical Journal* 633: 465–473.

Saumon D, Chabrier G, and van Horn HM (1995) An equation of state for low-mass stars and giant planets. *Astrophysical Journal* 99(supplement): 713.

Saumon D and Guillot T (2004) Shock compression of deuterium and the interiors of Jupiter and Saturn. *Astrophysical Journal* 609: 1170–1180.
Saumon D, Hubbard WB, Burrows A, Guillot T, Lunine JI, and Chabrier G (1996) A theory of extrasolar giant planets. *Astrophysical Journal* 460: 993.
Schneider J, Dedieu C, Le Sidaner P, Savalle R, and Zolotukhin I (2011) Defining and cataloging exoplanets: The exoplanet.eu database. *Astronomy and Astrophysics* 532: A79.
Schouten JA, de Kuijper A, and Michels JPJ (1991) Critical line of He-H_2 up to 2500 K and the influence of attraction on fluid-fluid separation. *Physical Review B* 44: 6630–6634.
Seager S and Deming D (2010) Exoplanet atmospheres. *Annual Review of Astronomy and Astrophysics* 48: 631–672.
Seager S and Sasselov DD (2000) Theoretical transmission spectra during extrasolar giant planet transits. *Astrophysical Journal* 537: 916–921.
Seiff A, Kirk DB, Knight TCD, et al. (1998) Thermal structure of Jupiter's atmosphere near the edge of a 5-μm hot spot in the north equatorial belt. *Journal of Geophysical Research* 103: 22857–22890.
Showman AP and Guillot T (2002) Atmospheric circulation and tides of "51 Pegasus b-like" planets. *Astronomy and Astrophysics* 385: 166–180.
Showman AP and Ingersoll AP (1998) Interpretation of Galileo probe data and implications for Jupiter's dry downdrafts. *Icarus* 132: 205–220.
Showman AP, Kaspi Y, and Flierl GR (2011) Scaling laws for convection and jet speeds in the giant planets. *Icarus* 211: 1258–1273.
Simon-Miller AA, Gierasch PJ, Beebe RF, et al. (2002) New observational results concerning Jupiter's Great Red Spot. *Icarus* 158: 249–266.
Sing DK, Désert J-M, Lecavelier Des Etangs A, et al. (2009) Transit spectrophotometry of the exoplanet HD 189733b. I. Searching for water but finding haze with HST NICMOS. *Astronomy and Astrophysics* 505: 891–899.
Snellen IAG, de Kok RJ, de Mooij EJW, and Albrecht S (2010) The orbital motion, absolute mass and high-altitude winds of exoplanet HD209458b. *Nature* 465: 1049–1051.
Soderlund KM, Heimpel MH, King EM, and Aurnou JM (2013) Turbulent models of ice giant internal dynamics: Dynamos, heat transfer, and zonal flows. *Icarus* 224: 97–113.
Spiegel DS and Burrows A (2013) Thermal processes governing hot-Jupiter radii. *Astrophysical Journal* 772: 76.
Spiegel DS, Silverio K, and Burrows A (2009) Can TiO explain thermal inversions in the upper atmospheres of irradiated giant planets? *Astrophysical Journal* 699: 1487–1500.
Spilker LJ (2012) Cassini: Science highlights from the equinox and solstice missions. In: *Lunar and Planetary Institute Science Conference Abstracts*, vol. 43, p. 1358.
Sromovsky LA, Fry PM, and Kim JH (2011) Methane on Uranus: The case for a compact CH_4 cloud layer at low latitudes and a severe CH_4 depletion at high-latitudes based on re-analysis of Voyager occultation measurements and STIS spectroscopy. *Icarus* 215: 292–312.
Stanley S and Bloxham J (2004) Convective-region geometry as the cause of Uranus' and Neptune's unusual magnetic fields. *Nature* 428: 151–153.
Stanley S and Glatzmaier GA (2010) Dynamo models for planets other than Earth. *Space Science Reviews* 152: 617–649.
Stevenson DJ (1982) Interiors of the giant planets. *Annual Review of Earth and Planetary Sciences* 10: 257–295.
Stevenson DJ (1983) Planetary magnetic fields. *Reports of Progress in Physics* 46: 555–557.
Stevenson DJ (1985) Cosmochemistry and structure of the giant planets and their satellites. *Icarus* 62: 4–15.
Stevenson DJ and Salpeter EE (1977a) The dynamics and helium distribution in hydrogen-helium fluid planets. *Astrophysical Journal* S35: 239–261.
Stevenson DJ and Salpeter EE (1977b) The phase diagram and transport properties for hydrogen-helium fluid planets. *Astrophysical Journal* S35: 221–237.
Sudarsky D, Burrows A, and Hubeny I (2003) Theoretical spectra and atmospheres of extrasolar giant planets. *Astrophysical Journal* 588: 1121–1148.
Trafton LM (1967) Model atmospheres of the major planets. *Astrophysical Journal* 147: 765–781.
Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Nature* 435: 459–461.
Vidal-Madjar A, Huitson CM, Lecavelier Des Etangs A, et al. (2011) The upper atmosphere of the exoplanet HD209458b revealed by the sodium D lines: Temperature-pressure profile, ionization layer and thermosphere (Corrigendum). *Astronomy and Astrophysics* 533: C4.
Vidal-Madjar A, Lecavelier des Etangs A, Désert J-M, et al. (2003) An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* 422: 143–146.
Visscher C and Moses JI (2011) Quenching of carbon monoxide and methane in the atmospheres of cool brown dwarfs and hot Jupiters. *Astrophysical Journal* 738: 72.
von Zahn U, Hunten DM, and Lehmacher G (1998) Helium in Jupiter's atmosphere: Results from the Galileo probe helium interferometer experiment. *Journal of Geophysical Research* 103: 22815–22830.
Vorberger J, Tamblyn I, Militzer B, and Bonev SA (2007) Hydrogen-helium mixtures in the interiors of giant planets. *Physical Review B* 75(2): 024206.
Vorontsov SV, Zharkov VN, and Lubimov VM (1976) The free oscillations of Jupiter and Saturn. *Icarus* 27: 109–118.
Wahl SM, Wilson HF, and Militzer B (2013) Solubility of iron in metallic hydrogen and stability of dense cores in giant planets. *Astrophysical Journal* 773: 95.
Warwick JW, Evans DR, Peltzer GR, et al. (1989) Voyager planetary radio astronomy at Neptune. *Science* 246: 1498–1501.
Warwick JW, Evans DR, Romig JH, et al. (1986) Voyager 2 radio observations of Uranus. *Science* 233: 102–106.
Weir ST, Mitchell AC, and Nellis WJ (1996) Metallization of fluid molecular hydrogen at 140 GPa (1.4 Mbar). *Physical Review Letters* 76: 1860–1863.
Wilson HF and Militzer B (2010) Sequestration of noble gases in giant planet interiors. *Physical Review Letters* 104(12): 121101.
Wilson HF and Militzer B (2012) Solubility of water ice in metallic hydrogen: Consequences for core erosion in gas giant planets. *Astrophysical Journal* 745: 54.
Winn JN and Holman MJ (2005) Obliquity tides on hot Jupiters. *Astrophysical Journal Letters* 628: L159–L162.
Wong MH, Mahaffy PR, Atreya SK, Niemann HB, and Owen TC (2004) Updated Galileo probe mass spectrometer measurements of carbon, oxygen, nitrogen, and sulfur on Jupiter. *Icarus* 171: 153–170.
Wright JT, Fakhouri O, Marcy GW, et al. (2011) The exoplanet orbit, database. *Publications of Astronomical Society of Pacific* 123: 412–422.
Youdin AN and Mitchell JL (2010) The mechanical greenhouse: Burial of heat by turbulence in hot Jupiter atmospheres. *Astrophysical Journal* 721: 1113–1126.
Zharkov VN and Trubitsyn VP (1974) Internal constitution and the figures of the giant planets. *Physics of the Earth and Planetary Interiors* 8: 105–107.
Zharkov VN and Trubitsyn VP (1978) *Physics of Planetary Interiors*. *Astronomy and Astrophysics Series* Tucson, AZ: Pachart.

This page intentionally left blank

10.17 The Origin of the Natural Satellites

SJ Peale, University of California, Santa Barbara, CA, USA
RM Canup, Southwest Research Institute, Boulder, CO, USA

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by S J Peale, Volume 10, pp. 465–508, © 2007, Elsevier B.V.

10.17.1	**Introduction**	559
10.17.2	**Earth–Moon System**	560
10.17.3	**Mars System**	567
10.17.4	**Jupiter System**	569
10.17.4.1	SEMM Model	570
10.17.4.2	Planetesimal Capture Model	572
10.17.4.3	Starved Accretion Disk Model	572
10.17.5	**Saturn System**	578
10.17.6	**Uranus System**	583
10.17.7	**Neptune System**	584
10.17.8	**Pluto System**	586
10.17.9	**Irregular Satellites**	588
Appendix A	**Accretion Disks**	589
Appendix B	**Tides**	594
Acknowledgments		598
References		598

10.17.1 Introduction

This chapter is an update to the chapter of the same title in the first edition of this volume of the *Treatise on Geophysics*. The information on, for example, the Mars and Neptune systems has not changed much, whereas new research and observations have motivated significant changes in the discussions of especially the Earth–Moon, Pluto, Saturn, and Uranus systems. This chapter will not be all inclusive, but we will attempt to select those thoughts and mechanisms that are currently prominent in the literature and will describe both virtues and caveats in the discussion. The authors' prejudices and judgments will no doubt be evident, but we hope advice from our colleagues has tempered the impact of such. The origin of the satellites constrains the origin of the solar system, and we hope this chapter will be a suitable introduction to researchers who seek to understand the origin of our solar system and that of planetary systems in general.

Each system of satellites is unique, and it is impossible to apply the same detailed sequence of events to account for the origin of the different systems. Of all the planetary satellites, our own Moon has generated the most curiosity about its origin and the subsequent tidal evolution that has placed constraints on the theories. The dynamic mechanisms that have been proposed for the formation of the Moon are evaluated by Boss and Peale (1986), and one or another of these has been proposed for other natural satellites. The mechanisms include what have been called rotational fission, precipitation fission, intact capture, disintegrative capture, binary accretion, and formation by giant impact. In rotational fission, favored by Darwin (1879, 1880), the Moon is derived from material shed from a rapidly spinning Earth as a result of a dynamic fission instability. Precipitation fission (Ringwood, 1970, 1972) requires a massive primordial atmosphere sufficiently hot to vaporize silicates. As the atmosphere cools, the silicates condense and precipitate to the Earth's equatorial plane where they can collect into the Moon, while the volatiles in the atmosphere are swept away by a T Tauri solar wind. For intact capture, the Moon is assumed to form from the accretion of planetesimals in heliocentric orbit relatively close to that of the Earth. During a close approach to the Earth, energy must be removed from the Moon to leave it in a bound orbit. Methods proposed for this energy loss include collision with an already existing satellite (Dyczmons, 1978), atmospheric drag (Horedt, 1976), an encounter with another object in heliocentric orbit when both the Moon and the other object are within the Earth's sphere of influence (Ruskol, 1972a), changes in the masses of the Sun (Szebehely and Evans, 1980) or the Earth (Lyttleton, 1967), and the dissipation of tidal energy during the time of close approach (Alfvén and Arrhenius, 1969, 1972, 1976; Conway, 1982; Gerstenkorn, 1955, 1969; Singer, 1968, 1970, 1972; Singer and Bandermann, 1970; Winters and Malcuit, 1977). Disintegrative capture involves the tidal disruption of a proto-Moon during a close encounter with the Earth, where some of the resulting fragments could be more easily captured into bound orbit to later accumulate into the Moon (Alfvén, 1963; Alfvén and Arrhenius, 1969; Öpik, 1969, 1971). In binary accretion, the Moon forms in the Earth's orbit from an accretion disk simultaneously with the formation of the Earth (Harris, 1977, 1978; Harris and Kaula, 1975; Ruskol, 1961, 1963, 1972a,b,c, 1975). A formation by giant impact takes advantage of the fact that the last stages of accretion of the terrestrial planets involved large bodies (Wetherill, 1985). The result of such an impact is likely to leave sufficient material beyond the Roche radius R_R to accumulate into the Moon (Cameron and Ward, 1976; Canup, 2004a,b; Hartmann and

Davis, 1975; Öpik, 1971). Only the giant impact origin of the Moon has survived rather intense scrutiny (see Boss and Peale, 1986 for a detailed rejection of the other schemes). The details of the Moon-forming impact are undergoing considerable study in an attempt to understand the observation that isotope ratios of several elements are identical on Earth and Moon. Whereas a giant impactor would generally have had an isotope distribution quite different from that of the Earth, the current Moon's surface and the Earth's mantle are made of the same stuff.

As formation of satellites in accretion disks is thought to be the dominant process for the regular satellites of the giant planets (regular satellites are those in nearly circular orbits coplanar with the planet's equatorial plane), we include a description of the processes and parameters appropriate for such accretion disks in **Appendix A**. In **Appendix B**, we outline a simple theory of gravitational tides and the dissipation therein that result in orbital and spin evolution. This evolution in many instances constrains the choice of processes and events involved in the origin of particular satellites. We do not include tables of orbital parameters or physical properties of the satellites, but refer the reader to the websites http://ssd.jpl.nasa.gov/?sat_elem and http://ssd.jpl.nasa.gov/?sat_phys_par for frequently updated total number of satellites with their orbital parameters at the first website and their updated physical properties in the second.

In **Section 10.17.2**, we develop the arguments supporting the giant impact origin of the Earth's Moon, which is necessary to account for the large angular momentum of the system along with a volatile and iron-poor moon. But in addition, we must account for the plethora of identical isotope ratios on the Earth and Moon. The equatorial orbits of Mars' satellites require their origin in a dissipative disk of debris, where means of accounting for a suspected grossly different composition of the satellites from Mars itself are discussed in **Section 10.17.3**. We describe recent attempts at understanding the origin of Jupiter's regular satellites by a generalization of binary accretion during the last stages of Jupiter's formation in **Section 10.17.4**. The constraints imposed by information obtained from the Galileo spacecraft mission guide these attempts, but all involve assumptions of various degrees of uncertainty. But considerable progress has been made recently, and a self-consistent and plausible theory of origin of Jupiter's regular satellites within the theory of the accretion of Jupiter itself is emerging.

In many ways, the Saturn system of regular satellites is much more complicated than Jupiter's, where nearly all of the mass in the system is in the single large satellite Titan, and small icy satellites and icy rings appear inside Titan's orbit where we might have expected large rocky satellites as in the Jupiter system. An additional complication is that the fraction of each satellite that is not ice is not a monotonic function of the orbital semimajor axis. The origin of Saturn's satellites is discussed in **Section 10.17.5**, where we indicate that this origin is among the least well understood in the solar system. The Uranian satellites are covered in **Section 10.17.6**, where the regular nature of these satellites in spite of a planetary obliquity of 97° implies that the disk from which these satellites formed was created in the giant impact that led to Uranus' large obliquity perhaps by disrupting an existing satellite system that was reassembled from the resulting equatorial disk. An alternative would be that the Uranian obliquity resulted from adiabatic processes that allowed the equatorial orbits of the satellites to be preserved, although such a process has not been found. Details of formation of the Uranian satellites in the collision scenario have been investigated but without detailed accretion processes. Neptune's satellite system, discussed in **Section 10.17.7**, is perhaps better understood than most. A plausible series of events centered around the essentially intact capture of the large retrograde satellite Triton account well for all of the observational properties of this system. The Pluto–Charon system (**Section 10.17.8**), like the Earth–Moon system, is characterized by a large specific angular momentum, which likely resulted from an oblique giant impact. The picture is complicated by the discovery of four additional small satellites coplanar with Charon's orbit, which are too distant to have been created simultaneously with Charon and their origin remains a mystery.

Minor satellites include the irregular satellites with large semimajor axes, eccentricities, and inclinations that orbit all of the major planets and the small close satellites often associated with rings of small particles. The former are distinguished by a capture origin (**Section 10.17.9**) and the latter by having been broken up and reassembled, probably several times, by collisions among themselves, with comets or with planetesimals passing through the planets' Hill spheres. The disintegrations of these close satellites are associated with supplying the small particles, which make up the observed rings.

Many asteroids and Kuiper Belt objects (KBOs) also have satellites, where the widely separated binaries probably result from capture (Goldreich et al., 2002) and those in close orbits most probably from collisions (Canup, 2005). But some small, close asteroid binaries may form from the rotational fission of a rubble pile, where the rotational acceleration comes from the thermal Yarkovsky-O'Keefe-Radzievskii-Paddock (YORP) effect (Walsh et al., 2008). Finally, Nesvorny et al. (2010) had proposed formation of the wide binaries of comparable mass components in the Kuiper Belt from gravitational collapse of a swarm of particles. The particles can be brought into the unstable swarms perhaps by assembly in turbulent eddies (Cuzzi et al., 2001, 2008; Johansen et al., 2006) or by a streaming instability (Johansen and Youdin, 2007; Johansen et al., 2007, 2009, 2011, 2012; Youdin and Johansen, 2007). We will not discuss the small body satellites beyond these assertions. Rotational fission, precipitation fission, and disintegrative capture among those failed processes offered for the formation of the Moon also fail to find any likely applications to the satellites of Mars, Pluto, and the major planets.

10.17.2 Earth–Moon System

As additional observational constraints are imposed, the ingenious theories of origin and evolution of the Moon become more complicated. Attempts to accommodate the constraints in modifications of the theories, albeit sometimes successful on the surface, introduce uncertainties and leave gaps in the evolution, which leave us without a completely definitive theory of origin. However, all theories keep the giant impact as the most likely beginning of the process leading to our current Earth–Moon configuration, and details described in this section may generate additional ideas.

The Moon's mass is 1.2% of the Earth's mass, which is the largest fractional mass of all the satellites in the solar system

with the exception of that in the Pluto–Charon system (if we ignore asteroid and Kuiper Belt satellites). In addition, the angular momentum of the system of $L_{EM} = 3.5 \times 10^{41}$ g cm^2 s^{-1} would lead to a 4 h rotation period of the Earth if the Moon were incorporated within. This large specific angular momentum is a key constraint on the origin. The expansion of the lunar orbit by tidal interaction with the Earth (Goldreich, 1966; Touma and Wisdom, 1994) is consistent with formation close to the Roche radius $R_R = 2.9R_E$, where R_E is the Earth's radius. Accretion of the Moon inside this distance would be prevented by tidal forces. At a density of 3.34 g cm^{-3}, the Moon is depleted in iron relative to the Earth, and samples of lunar material are depleted in volatiles with vaporization temperatures < 1300 K in solar nebula conditions (e.g., Jones and Palme, 2000). A thick layer of anorthositic crust whose constituents have a positive Eu anomaly is complemented with a Eu deficiency in the mare basalts. The thickness of the anorthositic crust coupled with the complementary Eu anomalies is consistent with the crystallization of the lighter anorthositic material in a few hundred kilometer deep magma ocean with the crystals floating to the top. The floating crystals are enriched in Eu leaving the magma, from which the mare basalts are later derived, deficient in this element (Taylor and Jakes, 1974; Wood, 1972; Wood et al., 1970). Limits on lunar contraction ($\lesssim 1$ km in radius) during cooling have been inferred from the lack of stress features on its surface, which has been interpreted as implying that the interior of the Moon was initially cold beneath the magma ocean (Solomon, 1986). While contraction of an initially hot Moon might be accommodated without the expected surface scarps (see the discussion in the succeeding text), the recent detection of subsurface gravity anomalies by the Gravity Recovery And Interior Laboratory (GRAIL) spacecraft also suggests an initially cool lunar interior (Andrews-Hanna et al., 2013) in contrast to an initially hot moon inferred from the giant impact origin.

With the exception of the volatiles, the lunar material and the Earth's mantle share a number of very close compositional similarities, where isotope ratios for several elements in lunar samples are identical to those on Earth within experimental uncertainty (Lugmair and Shukolyukov, 1998; Touboul et al., 2007; Wiechert et al., 2001; Zhang et al., 2012). The current angular momentum and the Moon's overall depletion in iron are consistent with the formation of the Moon by the impact of a differentiated Mars-sized body with a differentiated Earth (Canup, 2004a), which we shall refer to as the canonical impact. However, it is unlikely that such an impact alone can account for the isotopic similarities between the Moon and the silicate Earth, which has led to a suggestion of mixing of the disk vapor with the Earth's silicate atmosphere after the impact (Pahlevan and Stevenson, 2007) as well as to the recent exploration of alternative impact scenarios (Canup, 2012; Ćuk and Stewart, 2012; Reufer et al., 2012).

The formation of the Moon by giant impact was first published by Hartmann and Davis (1975) with a lunar-sized impactor that could account for the iron and volatile deficits, but Cameron and Ward (1976) realized that the grazing collision of a Mars-sized body could also account for the large angular momentum of the Earth–Moon system. Models of terrestrial planet accretion (e.g., Wetherill, 1985, 1992) showed that the Earth was likely to have undergone such Mars-sized impacts late in its accretion history and it is this scenario that has become the paradigm for the formation of the Moon. (For a delightful account of the development of the idea of giant the impact origin of the Moon by the four authors, see MacKenzie (2003), reviewed by Burns (2003).)

Hafnium–tungsten chronology has been used to estimate the time of the solidification of the lunar magma ocean. Radioactive ^{182}Hf decays to ^{182}W with a half-life $\tau_{1/2} = 9$ million years. Both elements are refractory, so the planets should have the same overall abundances as the solar system as a whole. The most recent analysis was limited to metals in returned lunar samples. This choice avoids augmentation of the ^{182}W from the capture of cosmic ray-generated neutrons by ^{181}Ta, since Ta is excluded from the metals. Any excess ^{182}W in the metals can thereby be attributed solely to production from ^{182}Hf decay. Recent experiments have detected no excess ^{182}W in metallic parts of 18 lunar samples, which implies that Hf was extinct at the time of solidification (Touboul et al., 2007). This result implies that solidification occurred more than 60 million years after the formation of the oldest constituents of meteorites, the calcium–aluminum-rich inclusions (CAIs). This is in contrast with earlier Hf–W chronology, which inferred a solidification age of 30–50 million years (Kleine et al., 2005), which was thought consistent with the timescale within which a giant impact was likely to have occurred. The late solidification of the Moon is supported by ^{207}Pb–^{206}Pb, ^{147}Sm–^{143}Nd, and ^{146}Sm–^{142}Nd chronologies performed on anorthositic material, which are consistent with crystallization as much as 200 million years after the formation of the CAIs (Borg et al., 2011). Since the solidification time of a magma ocean is ~10 million years (Elkins-Tanton et al., 2011), the giant impact may have occurred much later than previously assumed, or a scheme different from crystals of anorthositic material floating to the surface of a magma ocean to account for the anorthositic crust must be invoked. Contradicting the inference of a late formation of the Moon is the work of König et al. (2011), who find evidence for an early formation of the Earth's core associated with an inferred concurrent formation of the Moon. But the Earth's core probably formed long before the impact. Using Lu–Hf chronology, Taylor et al. (2009) argued for a rapid lunar differentiation implying an early formation. Still, Touboul et al. finding no excess ^{182}W in lunar metals is a rather clean result favoring a formation of the Moon after ^{182}Hf is extinct.

The angular momentum relative to the center of mass of the target (mass M_t) delivered by an impactor of mass $M_i = \gamma M_T$, where $M_T = M_t + M_i$ is the total mass, is (Canup, 2004a)

$$L_i = bM_T^{5/3} f(\gamma) \sqrt{\frac{2G}{\left(4\pi\rho_p/3\right)^{1/3}}} \left(\frac{v_i}{v_{esc}}\right) \approx 1.3 L_{EM} b \left(\frac{M_T}{M_E}\right)^{5/3} \left(\frac{\gamma}{0.1}\right) \left(\frac{v_i}{v_{esc}}\right) \quad [1]$$

where $b = \sin\xi$ is the impact parameter divided by $R_t + R_i$, the sum of the target and impactor radii; ξ is the angle between the line connecting the centers of M_t and M_i and the impact trajectory at the time of first contact, such that grazing incidence has $\xi = 90°$ or $b = 1$; G is the gravitational constant; ρ_p is the common density of M_t and M_i; v_i is the velocity of the impactor; $v_{esc} = \sqrt{2GM_T/(R_t + R_i)}$ is the mutual escape velocity, $f(\gamma) = \gamma(1-\gamma)\sqrt{\gamma^{1/3} + (1-\gamma)^{1/3}}$; and M_E is the Earth's mass. The complication of eqn [1] comes from introducing v_{esc} into the equation for angular momentum.

For the simplest case in which a single impact provides all the Earth–Moon system angular momentum, L_i must be comparable to or greater than that in the current Earth and Moon (L_{EM}). Some angular momentum will be lost in escaping material both from the initial impact and during the assembly of the Moon from the impact-generated disk. The magnitude of the former depends primarily on the impact velocity, while models of the Moon's accretion suggest that the latter would have been of order of a few $\times 10^{-2}\ L_{EM}$ (Kokubo et al., 2000; Salmon and Canup, 2012). The Earth–Moon angular momentum would have also been decreased by a few percent by tides raised on the Earth by the Sun that slow the Earth's rotation (e.g., Canup et al., 1999). For a given value of L_i, there is a multidimensional parameter space involving b, M_T, γ, and v_i that must be explored to find successful candidate impacts. Accommodation of values of $L_i \approx 2L_{EM}$ in more energetic impacts (Canup, 2012; Ćuk and Stewart, 2012) will be discussed later.

A successful lunar-forming impact produces a disk capable of later yielding an iron-depleted moon outside the Roche limit. This condition is met either when a lunar mass or more is placed into orbits outside R_R by the impact itself or when the disk has sufficient mass and angular momentum so that viscous spreading after the impact can deliver sufficient mass beyond R_R. Simulations of the disk's evolution typically find that a single moon accretes just outside R_R, with inner disk material driven into the planet and $<5\%$ of the disk's mass escaping from the Earth's orbit (Crida and Charnoz, 2012; Ida et al., 1997; Kokubo et al., 2000). A mass and angular momentum conservation argument can be used to estimate the mass of the moon (M_M) that would accumulate from a disk as a function of the initial disk mass (M_D), the initial disk angular momentum (L_D), the initial position of the moon ($a = \lambda R_R$), and the angular momentum of material that escapes or collides with the Earth. Based on results of disk simulations and with $\lambda = 1.3$, the predicted moon mass is

$$\frac{M_M}{M_D} \approx 1.9\left(\frac{L_D}{M_D\sqrt{GM_E R_R}}\right) - 1.1 - 1.9\left(\frac{M_{esc}}{M_D}\right) \qquad [2]$$

where M_{esc} is the escaping mass (Ida et al., 1997; Kokubo et al., 2000). This equation provides a simple way to determine whether a disk produced by a hydrodynamic simulation could later yield a lunar mass moon. It is physically invalid for cases that would give $M_M/M_D > 1$; these correspond to initial disks with specific angular momenta too high to be consistent with the assumption that the final moon forms at $1.3R_R$.

Smooth particle hydrodynamics (SPH) is the method most used in numerical exploration of the impact parameter space. SPH is a Lagrangian method that describes the colliding objects by a multitude of particles that are evolved due to self-gravity and pressure forces. Over time, the spatial resolution of such calculations has increased from $\sim 10^3$ particle simulations in Benz et al. (1986, 1987) to 10^5–10^6 particle simulations in recent works (e.g., Canup, 2004b, 2008, 2012; Canup and Asphaug, 2001; Ćuk and Stewart, 2012; Reufer et al., 2012). At the same time, there have been algorithm and equation of state (Melosh, 2007; see also the discussion in Canup, 2004b) improvements that have influenced the results. Alternatively, impacts can be modeled with a grid-based, Eulerian hydrocode (Canup et al., 2013; Wada et al., 2006). Simulations with SPH and the grid-code CTH yield broadly similar results, with the predicted protolunar disk mass and its angular momentum converging to typically within about 10% for simulations of very similar impact conditions (Canup et al., 2013).

SPH simulations show that for $L_i \approx L_{EM}$, those impacts that produce a relatively iron-poor disk having M_M (from eqn [2]) of at least a lunar mass typically have $b \sim 0.7$ ($\sim 45°$ impact angle), a low impact velocity $v_i/v_{esc} < 1.1$, and an impactor mass ratio $0.1 < \gamma < 0.15$ (a roughly Mars-sized impactor) for $M_T \sim M_E$ (Canup, 2004a; Canup and Asphaug, 2001). An impact involving a low-velocity, Mars-sized impactor with $L_i \approx L_{EM}$ has generally been considered the 'canonical' impact for forming the Moon, because it accounts for the main dynamical properties of the Earth and Moon and the Moon's relative lack of iron while relying on conditions that are probable during the late stages of terrestrial planet accretion, that is, large, randomly oriented impacts, with 45° being the most probable impact angle (e.g., Agnor et al., 1999). However, because such cases yield a disk formed primarily from material originating from the impactor, the observed identical isotope ratios for lunar and Earth mantle materials cannot prevail.

It has been known for decades that the Earth and Moon share identical oxygen isotope compositions that are distinct from those of most meteorites and Mars (e.g., Wiechert et al., 2001). This would be consistent with the Moon's composition if the Moon formed directly from the Earth's mantle. If instead the disk (and the Moon) formed primarily from impactor material, an impactor with an identical oxygen composition as the Earth would then be required. Because oxygen compositions vary with heliocentric position, it was suggested that this similarity would be expected if the impactor formed near the Earth (e.g., Belbruno and Gott, 2005; Wiechert et al., 2001). This seemed to agree well with the canonical impact, which required a low impact velocity and therefore an orbit close to that of the Earth. However, this explanation was brought into serious question by Pahlevan and Stevenson (2007), who argued that it is very improbable that the impactor and Earth would share the same oxygen composition even if they were in close orbits, given the degree of radial mixing predicted by planet accretion models. Using the impacts predicted in one of Chambers' (2001) terrestrial accretion simulations, together with the assumption that the oxygen composition of planetary embryos varied linearly with initial heliocentric distance, they estimated that the typical difference between the composition of an impactor and its final target planet would have been comparable to the current difference between Earth and Mars, which is about 50 times greater than the allowable difference in the Earth–Moon oxygen compositions. A moon accreted from a disk composed primarily of impactor-derived material would then be expected to have a substantially different oxygen composition than the Earth. Furthermore, the isotopic composition of multiple other elements now also appears identical in the silicate Earth and Moon, including chromium (Lugmair and Shukolyukov, 1998), titanium (Zhang et al., 2012), tungsten (Touboul et al., 2007), and silicon (Georg et al., 2007). Georg et al. pointed out that Si partition into the core is a high-pressure process that depletes the heavy isotopes more than the light ones and that requires at least 0.15 Earth masses to become

effective. That means that even if the oxygen, chromium, and titanium isotopic compositions were identical in the Earth's mantle and the Mars-sized impactor, the silicon composition of the impactor would likely not be because of partitioning of the heavier Si isotopes into the Earth's core. This supports the Pahlevan–Stevenson assertion that Earth and impactor could not have identical isotopic compositions before the impact.

As a solution, Pahlevan and Stevenson (2007) proposed that the vapor components of the disk and the Earth's silicate atmosphere diffusively mixed after the impact via vigorous convection, allowing the disk to compositionally equilibrate with the Earth before the Moon forms. However, as stressed by Melosh (2009), for this to occur, one must assume that the convection mechanism does not also transport angular momentum at the same efficiency, because if it did, the disk would spread into the Earth on a similar timescale and diffusion from the Earth's atmosphere would have to operate upstream against the flow of this incoming material. Still, this may be a valid assumption, since thermal convection is now thought to be inefficient in transporting angular momentum in disks (Stone et al., 2000). Equilibration is estimated to require $\sim 10^2$–10^3 years, so that the Moon's accumulation must be delayed by this amount of time after the impact (Pahlevan and Stevenson, 2007). The mixing must involve the whole of the disk to equalize the planet-disk compositions, which may be difficult if a substantial fraction of the Moon's mass is initially outside the Roche limit (e.g., Salmon and Canup, 2012). It is not clear that vapor equilibrium of refractory elements in a canonical impact (including notably titanium) is possible, although if the disk lasts for $\geq 10^2$ years, there may be sufficient time for even refractory elements in the magma layer to equilibrate with the disk's vapor atmosphere (Zhang et al., 2012). In addition, the titanium isotope composition of Mars appears much closer to the Moon's composition than is true for oxygen. Within current uncertainties, the Mars–Earth Ti difference may only be about a factor of 2 larger than the Moon–Earth difference (e.g., Canup, 2012, supplemental material), so that a disk composed primarily of material from a Mars-composition impactor might not require much equilibration of Ti to be consistent with the current Moon. Thus, titanium does not provide as strong a constraint on equilibration as oxygen, for which the Earth–Mars difference is much larger.

Such issues motivated additional impact studies with L_i close to L_{EM} with varying assumptions to see if a disk with planet like composition could be produced. Comparisons between simulations performed with SPH versus CTH found no systematic influence of simulation method on predicted disk composition (Canup et al., 2013). To quantify the compositional difference between the silicate portions of the disk and planet, one can define a deviation percentage (after Reufer et al., 2012),

$$\delta f_T = \left[\frac{F_{D,tar}}{F_{P,tar} - 1}\right] \times 100 \qquad [3]$$

where $F_{D,tar}$ and $F_{P,tar}$ are the mass fractions of the silicate portions of the disk and the planet derived from the target's mantle. Identical disk-planet compositions have $\delta f_T = 0$, whereas a disk that contains fractionally more (less) impactor-derived silicate than the final planet has $\delta f_T < 0$ ($\delta f_T > 0$). For a Mars-like composition impactor, consistency with the Earth–Moon oxygen composition requires $|\delta f_T|$ less than a few percent (e.g., Canup, 2012).

Prior works had assumed a target and impactor that were not rotating prior to their collision, an almost certainly incorrect assumption. Canup (2008) examined the effect of pre-impact rotation in impacts that leave a planet-disk system with an angular momentum, L, comparable to L_{EM}. She found that a target proto-Earth with a pre-impact retrograde rotation (i.e., in the opposite rotational sense as the impact itself) allowed for somewhat larger impactors and produced disks containing up to 40% target material. Reufer et al. (2012) considered higher-velocity, more head-on impacts in which a substantial fraction of the impact angular momentum escapes, so that L_i may be much greater than L_{EM} and still leave a planet-disk system with $L \sim L_{EM}$. They found that collisions with $1.2 \leq v_i/v_{esc} \leq 1.3$, $0.15 \leq \gamma \leq 0.2$, and $\xi = 30$–$40°$ produced appropriately massive disks that contained up to $\sim$60% target material.

The canonical impacts typically have $-90\% \leq \delta f_T \leq -70\%$, whereas the successful cases (producing a disk leading to a $\sim 1 M_M$ Moon) in Canup (2008) have $\delta f_T \leq -50\%$ and those in Reufer et al. (2012) have $-40\% \leq \delta f_T \leq -30\%$. The more target-rich disks produced by Canup (2008) and Reufer et al. (2012) have values of $|\delta f_T|$ far greater than the required few % that would account for the Earth–Moon oxygen isotope similarity.

These failures to obtain a disk made up almost entirely of Earth mantle material have motivated proposals for additional modifications of the impact scenario, where the angular momentum of the resulting system is no longer constrained to be near L_{EM}. Ćuk and Stewart (2012) chose high-velocity, equatorial plane impacts ($v_i = 15$–25 km s^{-1}, 1.5–3 times v_{esc}) onto rapidly spinning Earth-mass targets ($M_t = M_E$) (periods 2.3–2.7 h corresponding to angular momenta of 1.9–$3.1 L_{EM}$), where $0.026 \leq M_i/M_t \leq 0.1$. The requirement of a near-fission rotation rate of the Earth prior to impact may be quite restrictive. Rapidly rotating planets are seen in N-body simulations of late-stage accretion (e.g., Agnor et al., 1999). However, such models tend to overestimate planetary rotation rates because they assume pure mergers and ignore the possible ejection of mass and angular momentum during large impacts. The most successful outcomes in the Ćuk–Stewart simulations, where the disk was sufficiently Earth mantle-dominated to satisfy the isotopic constraints (with $|\delta f_T|$ $\sim$several percent) and the predicted mass of the accreted satellite was near M_M, occurred for head-on or slightly retrograde impacts onto the higher angular momentum initial targets with impact velocities ~ 1.5–$2.5 v_{esc}$. In the successful Ćuk–Stewart impacts, the Earth rotation period is reduced, but the total remaining angular momentum is still $\gtrsim 2 L_{EM}$.

Canup (2012) reported simulations successful in producing a disk with the same composition of the Earth's mantle by colliding nonrotating targets and impactors of similar mass ($M_i/M_T \geq 0.4$) at low velocity. Here, the impact parameter was varied over a wide range yielding $0.35 \leq b \leq 0.7$, which determined the angular momentum of the Earth + disk and yielded a disk of sufficient mass to produce a Moon mass satellite. The silicate mantles of target and impactor were well mixed by the impact, so the resulting Earth mantle and disk had essentially the same silicate composition to satisfy the isotope constraints

(Nakajima and Stevenson, 2013). The most successful Canup examples yielded $\delta f_T = -0.8\%$ and -0.3%, predicted satellite masses of 1.64 and $1.09 M_M$, but left the Earth spinning with period of about 2 h with total angular momenta of 2.45 and $2.22 L_{EM}$.

The kinetic energy of both the new impact scenarios is comparable (although the Canup model involves a much larger impactor, the Cuk and Stewart model involves a higher impact velocity) and is about a factor of three larger than that of the canonical impacts. Such violent events may have difficulty explaining the recent inference that the Earth's mantle may not have been thoroughly mixed (Touboul et al., 2012), although Ćuk and Stewart (2012) argued that at least in the case of their impacts, this may result from nonuniform heating during the collision. In general, relaxing the angular momentum constraint in two different giant impact scenarios has allowed disk and Earth mantle to have compositions that are sufficiently alike in both the Ćuk and Stewart (2012) analysis and that of Canup (2012) to satisfy the isotope constraints in oxygen, titanium, and chromium (see succeeding text for the discussion on tungsten and silicon), but both leave the system with excess angular momentum that must be eliminated.

Ćuk and Stewart (2012) offered an ingenious scheme for reducing the system angular momentum. As the Moon recedes from the Earth due to torques from tides raised on the Earth, it is captured into the evection resonance with the Sun. The evection resonance occurs when the precession period of the Moon's perigee is commensurate with the Earth's orbital period about the Sun. For an Earth–Moon system with an angular momentum $\sim L_{EM}$, the resonance occurs when the Moon's semimajor axis is $a \approx 4.6$ Earth radii. This resonance had been previously investigated by Touma and Wisdom (1998), who found that it could play a key role in setting up capture into a resonance involving the lunar inclination that could explain the Moon's initial orbital inclination (see succeeding text). For the more rapidly spinning Earth (~ 2.5 h day), corresponding to a system angular momentum ~ 2–$2.5 L_{EM}$, the Earth's increased oblateness causes the location of the resonance to shift outward to $a \approx 7$ Earth radii. At this larger distance, the rate of the Moon's orbital expansion due to tidal interaction with the Earth is slower, which increases the likelihood of capture into evection. In the evection resonance, the line of apsides of the lunar orbit remains perpendicular to Earth–Sun direction. Angular momentum is tidally transferred to the Moon orbit from the Earth and from there to the Earth–Moon orbit about the Sun because of resonance. The orbital eccentricity increases to a value near 0.6 for $A_Q \approx 1$ (see eqn. 4) as the Moon's orbit expands until the expansion of the Moon's orbit stalls due to a balance between Earth tides (which cause a to expand) and satellite tides (which cause a to contract for high-eccentricity orbits). As long as the evection resonance is maintained, angular momentum of the Earth–Moon system is being reduced, and in the Ćuk and Stewart simulations, the Moon's prolonged residency in the resonance removes the excess angular momentum until the resonance is broken and the Moon commences its tidal evolution to its current distance from the Earth. However, there are some caveats in this scheme for removing the excess angular momentum.

A measure of the relative tidal dissipation in the Moon and Earth is given by the ratio

$$A_Q = \frac{k_{2M}}{k_{2E}} \frac{Q_E}{Q_M} \frac{M_E^2}{M_M^2} \frac{R_M^5}{R_E^5} \qquad [4]$$

where the subscript Q indicates the constant Q tidal model. The successful cases of Ćuk and Stewart, where systems with $L \approx L_{EM}$ were obtained, occurred only for a very narrow range of A_Q near 1. If lunar tides are too strong relative to Earth tides, the system leaves the resonance before sufficient angular momentum is removed, whereas if the lunar tides are too weak compared to the Earth tides, the lunar orbital eccentricity reaches overly high values that destabilize the resonance (Ćuk and Stewart, 2012). Both the Earth and Moon are likely to be nearly all molten initially with values of k_2 close to the fluid values, and $A \sim 1$ implies $Q_M \approx 10 Q_E$ (little dissipation in the Moon) from eqn [4]. There is no reason for Q_M to be so much larger than Q_E, but see the succeeding text.

The reader should be warned that the results of the calculations depend on the tidal model. Ward and Canup (2013), using the constant Δt ($Q \propto 1/\omega$) described in **Appendix B**, were also in some cases able to remove all the excess angular momentum while the Moon is trapped in the evection resonance. In their case, the measure of dissipation in the Moon to that in the Earth is given by

$$A_{\Delta t} = \frac{k_{2M}}{k_{2E}} \frac{\Delta t_M}{\Delta t_E} \left(\frac{M_E}{M_M}\right)^2 \left(\frac{R_M}{R_E}\right)^5 \qquad [5]$$

and a value of $A_{\Delta t} \approx 9$ is used in a successful case that removes nearly all of the excess angular momentum. Recall for this tidal model that $Q = 1/\omega \Delta t$ (**Appendix B**), so we could replace $k_2 \Delta t$ in eqn [5] with $k_2/Q\omega$, where Q is now that value appropriate for the frequency ω. The dominant tidal frequency in the Moon is close to the mean orbital motion n, whereas that in the Earth is the much larger $\dot{\psi} - n$, where $\dot{\psi}$ is the angular velocity of the Earth. Then, $k_{2M}\Delta t_M/(k_{2E}\Delta t_E) = (k_{2M}/k_{2E})(Q_E/Q_M)\left(\left(\dot{\psi} - n\right)/n\right)$, where again the Qs are those appropriate for the frequencies they multiply. But for constant Δt, $Q_E(\omega) = Q_E(n)(n/\omega)$, so $A_{\Delta t} \approx 10 Q_E(n)/Q_M(n)$ if the k_2s are comparable. The Ward–Canup result with $A = 9$ would then require the Qs for the Moon and Earth evaluated at the frequency n to be comparable, which contrasts with the constant Q model. Another reservation is that the large increase in the eccentricity during the angular momentum transfer could lead to a drastic increase in the lunar dissipation and the value of A_Q would drift far outside the narrow range for stability of the evection resonance (Wisdom, private communication). There may be a similar restriction on $A_{\Delta t}$. We should therefore regard both the Ćuk–Stewart and the Ward–Canup results as indicative that removal of the angular momentum in the evection resonance could be possible with the reservation that at least the constant Q results have to be relatively finely tuned, in which tuning might be frustrated by the increased dissipation in the Moon during the high-eccentricity phase, and that neither tidal model represents the behavior of real materials (e.g., Castillo-Rogez et al., 2011).

There are thus two viable schemes for creating a moon with identical isotopes to those of the Earth's mantle that are consistent with a giant impact origin of the Moon (Canup, 2012; Ćuk and Stewart, 2012) without requiring postimpact equilibration. Although both require a system angular momentum close to twice L_{EM}, at least a process has been demonstrated that could reduce the angular momentum to a value close to

L_{EM}, but not without important caveats. What is less clear is whether these models can account for the Earth–Moon similarities involving elements that partition strongly into metals and therefore can be affected by core formation, including tungsten and silicon. Whether the disk is formed from the target's mantle (as in the Ćuk and Stewart impacts) or from an approximately equal mix of target and impactor mantles (as in the Canup impacts), the core of the impactor is absorbed by the planet, while the disk contains little iron. Mixing and equilibration of the impactor's core with the target's mantle would tend to alter the planet's tungsten and silicon compositions relative to that of the disk (and the resulting Moon). However, it is possible that rapid merging of the impactor's large core with the Earth's core precludes efficient silicate–metal equilibration, because the latter requires emulsification down to centimeter scales (Dahl and Stevenson, 2010; Rubie et al., 2003).

Although the overall compositional and dynamical properties of the current Earth–Moon system appear consistent with these latest versions of giant impact origin, the transition from the distribution of material following the impact to the accreted Moon a little outside the Roche radius R_R that is also consistent with current observations remains less well understood. An impact-produced disk contains silicate magma and vapor, with initially 10–30% vapor by mass for canonical impacts and up to 90% vapor by mass for the high angular momentum impacts. Much or perhaps most of the disk mass initially orbits within the Roche limit. Interior to R_R, clumps among the magma formed by local gravitational instability are sheared apart by planetary tidal forces (Ward and Cameron, 1978). This generates a large viscosity that dissipates heat and causes the Roche interior disk to spread inward onto the planet and outward past R_R on a timescale τ_v. Material from the inner disk that spreads beyond R_R, together with material that was placed into orbits outside R_R by the impact itself, provides the material that can accrete into the Moon.

Early estimates that assumed a pure magma disk (no vapor) suggested $\tau_v \sim O(1)$ year (Ward and Cameron, 1978). This timescale is consistent with that seen in N-body lunar accretion simulations of a vapor-free disk of particles that find that the Moon accretes in less than a year (Crida and Charnoz, 2012; Ida et al., 1997; Kokubo et al., 2000). However, there is an inconsistency in modeling the protolunar disk as vapor-free, because the energy dissipated as a disk of molten or solid particles viscously spreads is sufficient to vaporize the disk. Vaporization renders the disk gravitationally stable unless the energy is radiated away. Silicate vapor and liquid phases coexist when the disk's photospheric temperature is of order ~2000 K. If the phases are vertically well mixed, the radiation rate of the disk at this temperature can limit the disk's spreading to a much longer timescale, $\tau_{rad} \sim O(10^2)$ years (Thompson and Stevenson, 1988). Thompson and Stevenson (1988) suggested that a lower viscosity consistent with this timescale could result as the disk hovers on the brink of local gravitational instability with a sound speed just below the critical value for marginal stability. The latter can occur in a two-phase disk if the gas mass fraction is very low – only a few percent – with the bulk of the material comprised of liquid droplets that are vertically well mixed with the vapor. An alternative, stratified disk structure can result if the liquid instead settles, forming a midplane magma layer surrounded by a vapor disk atmosphere (Ward, 2012). In either case, the lifetime of the Roche interior disk is $\geq 10^2$ years.

The $\leq$1-year accretion timescale predicted by N-body lunar accretion models is at odds with these much longer disk lifetimes. More recent simulations that pair an N-body model for material outside the Roche limit with a fluid model for an inner disk that persists for $\sim 10^2$ years find that the Moon takes $\sim 10^2$ years to complete its accretion (Salmon and Canup, 2012), which may provide sufficient time for the equilibration of oxygen (Pahlevan and Stevenson, 2007). Accretion of the Moon in $\leq$1 year implies a fully molten Moon, whereas accretion over 10^2 years may allow for partial cooling for favorably chosen conditions (e.g., Pritchard and Stevenson, 2000). A molten Moon is not consistent with the most straightforward interpretation of the lack of global-scale thrust faults on the Moon, which implies limited contraction and an initially cool interior (Solomon, 1986). There are uncertainties in the properties of the lunar crust and in the thermal history that seem to allow ranges of parameters and processes that could accommodate an initially molten Moon in spite of the lack of observable indications of a significant radius change (Pritchard and Stevenson, 2000). But recently, the GRAIL spacecraft has identified a system of subsurface gravity anomalies hundreds of kilometers in length that are interpreted as dikes formed during the expansion of the early lithosphere (Andrews-Hanna et al., 2013). Such an early period of expansion is predicted for an initial Moon with a cool interior and an overlying magma ocean as the interior warms during the Moon's first billion years (Solomon, 1977). The GRAIL data thus provide additional evidence for an initially cool lunar interior. Any surface indication of the expansion could have been erased by the late heavy bombardment (LHB) on the Moon, where the existence of the dikes seems to imply that the LHB was not so intense as to create a second magma ocean. But an initially cold Moon is not consistent with the results of the giant impact.

The most likely consequence of the accretion of the Moon from a disk of material generated by a giant impact is a hot, at least partially molten Moon orbiting just outside R_R in the Earth's equatorial plane. Before we leave the Earth–Moon system, we must then account for the Moon's current orbital properties under the assumption that the semimajor axis of the Moon's orbit was increased from its initial to the current value by tidal torques. By integrating backward in time from the current state, Goldreich (1966) found that the Moon's orbit had an inclination to the Earth's equator of about 10° when the lunar semimajor axis was approximately $10R_E$. This result was confirmed by Touma and Wisdom (1994). The 10° inclination is not consistent with the Moon having formed in the Earth's equatorial plane from the debris generated by a giant impact. Moving the Moon from an equatorial orbit to one inclined by 10° is called the 'inclination problem.'

There are four proposals for accomplishing this bit of dynamics. If the Earth had a significant spin before impact like the Ćuk and Stewart initial condition, and a substantial fraction of the Moon remained in a coherent clump after the impact, the latter's orbit could have been inclined to the equatorial plane and subsequent accretion of the remaining lunar mass could leave it inclined (Canup, 2004a). But the equilibration of the isotope ratios would have been frustrated, even if the unlikely event occurred. Alternatively, an additional impact

on the Earth after the Moon-forming event could have altered the alignment of the Earth's spin. But this event would have had to have occurred while the Moon was very close (Ward, 2002), and it is not clear how the Moon would not have been contaminated with additional material with different isotope ratios – even if the first impact established equal ratios. The remaining two hypotheses do not obviously frustrate the identical isotope condition, but both still have caveats. They both involve resonant interactions to increase the initial Moon's orbital inclination – one with orbital resonances between the Moon's motion and the Sun (Touma and Wisdom, 1998) and the other with resonances between the Moon's motion and a remnant disk inside R_R (Ward and Canup, 2000).

In an integration that produces the current configuration when the Moon has reached its current distance from the Earth, Touma and Wisdom (1998) started the Moon in the equatorial plane of the Earth at a separation of $3.5R_E$ with an eccentricity of 0.01. The initial obliquity of the Earth is 10° and the initial Earth rotation period is 5 h. There is no circumplanetary disk present. Realistic rates of tidal evolution are used in the symplectic integrations that include the entire chaotic solar system, and dissipation in the Moon is included at a variety of dissipation rates. A tidal model was used employing the constant time lag as discussed in **Appendix B** but with the Mignard (1981a) formulation. The evection resonance is encountered when the Moon is at about $4.6R_E$, where the period of the periapse motion of the lunar orbit relative to an inertial reference is near one year. This resonance is called the evection resonance because the same term in the disturbing function gives rise to the 1.3° amplitude, 31.8 day periodic variation in the Moon's mean longitude called the evection. Capture into the evection resonance is certain if the eccentricity is below about 0.07 as the resonance is approached and if the rate of tidal evolution is sufficiently slow. With the assumed parameters, capture occurs and the eccentricity grows rapidly to large values, where the maximum value reached before the system escapes the resonance is determined by the value of $A_{\Delta t}$ defined by eqn [5]. For $A_{\Delta t}=0$ (no dissipation in the Moon), the maximum eccentricity is about 0.5 before escape and the eccentricity continues to climb after escape from the resonance because of tides raised on Earth and no dissipation in the Moon (Goldreich, 1963). For $A_{\Delta t}=10$ (high dissipation in the Moon), the maximum eccentricity is only about 0.15. For $1 \le A_{\Delta t} \le 10$, the energy dissipated in the Moon in about 8000 years is in the range $\sim 2\times10^{35}$–1.5×10^{36} ergs, which could lead to substantial melting (Touma and Wisdom, 1998).

After escape from the evection resonance, the continued expansion of the orbit further decreases the prograde motion of the orbit periapse and twice the time derivative of the evection resonance variable plus the retrograde motion of the lunar orbit node approaches zero near $6R_E$. The term in the Hamiltonian corresponding to this resonance has $e_M^2 i_M$ in the coefficient, but this resonance affects the inclination more than the eccentricity. Touma and Wisdom named this resonance the eviction – changing the e in evection to i to emphasize the inclination and noting that this resonance 'evicts' the Moon from an equatorial orbit. If A is not too large, the eviction resonance is approached with high eccentricity in the wrong direction for capture. Passage through the resonance leaves the eccentricity large and excites an inclination of 2°–3°. If A were now to increase drastically – perhaps because the continued high eccentricity has partially melted the interior – the dissipation in the Moon becomes so high that the semimajor axis is decreased as the eccentricity is reduced. This takes the system through the eviction resonance from the other direction where capture and subsequent evolution force the inclination to values between 9° and 13°. Escape from the resonance is effected because of the continued decrease in e_M, but the remnant inclination is preserved. Subsequent evolution brings the Moon to the current configuration. The Touma and Wisdom analysis was the first that allowed the Moon to evolve from an equatorial orbit to its current configuration from the effects of tidal dissipation alone.

An important constraint on the Touma–Wisdom model is that the Moon needs to begin its tidal evolution cold, which is contrary to our conclusion earlier that the Moon most likely started hot in at least a partially molten state (e.g., Pritchard and Stevenson, 2000). In addition, a more rapidly rotating initial Earth as advocated in Ćuk and Stewart (2012) moves the position of the evection outward, so that it takes longer for the Moon to encounter the resonance. However, from the expression for the tidal variation in the mean motion (Peale, 1988)

$$\frac{dn}{dt} = -\frac{9}{2}k_2\frac{M_M}{M_E}\frac{R_E^5 n^{16/3}}{(GM_E)^{5/3}}\frac{1}{Q_E} \qquad [6]$$

the timescale for expanding the lunar orbit from $3R_E$ to $7R_E$ is only $\sim 32Q_E$ years, where Q_E is likely to be < 10 for a partially molten Earth. The timescale for solidifying a magma ocean on the Moon is 1000 years for the first 80%, appropriate to maintaining melt on the surface from foundering of the cooled material, and 10 million years for the remainder appropriate for the conductive timescale of an anorthositic stagnant lid (Elkins-Tanton et al., 2011). The further distance of the evection resonance for the rapidly spinning Earth does not allow sufficient time for significant cooling of the Moon. The probable hot Moon remains a caveat in the Touma–Wisdom solution to the inclination problem. But if the inference from the GRAIL data that the Moon was initially cold can be supported by a believable evolution of the system, perhaps the Touma–Wisdom ideas can gain credibility.

The scheme involving disk interactions to change the Moon's inclination (Ward and Canup, 2000) depends on the coexistence of the Moon with a 0.5–1 lunar mass Roche interior disk for 10–100 years. Waves in the disk are generated at locations where the ratio of the local mean motion to that of the Moon involves two small integers. Torques that result from the gravitational interaction of the Moon with the distribution of mass in the wave structures can affect the lunar orbit. Such mean motion resonances (MMRs) from planar interactions are called Lindblad resonances, and they generate spiral density waves. Spiral bending waves are associated with the inclination of the Moon's orbit, where material in the disk is lifted out of the plane (e.g., Shu, 1984). When the Moon is just outside R_R, there are many MMRs in a disk that extends nearly to R_R. Torques from the waves generated in the disk force the Moon to larger semimajor axes, while the back reaction reduces the angular momentum contained in the disk. The resonances thereby move out through the disk and eventually leave its outer edge until the last strong one to leave the disk is the 3:1 inner vertical resonance (IVR) when the Moon's semimajor axis is a little over twice the disk radius.

The effect of the torque $T_{IVR} \propto \sigma_g \sin^2 I$ from the so-called bending waves, where I is the orbital inclination for the Moon and σ_g is the surface mass density of the inner disk, in addition to contributing to the growth of the semimajor axis a of the Moon's orbit, is an increase in the inclination at the rate (Ward and Canup, 2000)

$$\frac{dI}{dt} = \frac{T_{IVR}}{M_M a^2 \Omega_K \sin I}\left(\frac{3}{2}\cos I - 1\right) \quad [7]$$

where M_M is the lunar mass and Ω_K is the Kepler mean motion. With a disk mass of $0.75 M_M$ and an initial $I = 1°$, inclinations reached 12.3° and 14.5° as the Moon nears $6R_E$ for initial disk spreading times of 37 and 50 years, respectively (Ward and Canup, 2000). The growth of I stops if either the resonance moves outside the disk or the disk is depleted. The disk interaction mechanism requires control of the evolution by the 3:1 vertical resonance. Other resonances, if still present, can lead to inclination damping (e.g., Borderies et al., 1984; Ward and Hahn, 1994). However, the Lindblad resonances invoked by Borderies et al. to damp the inclination are out of the disk when the 3:1 vertical resonance is near the disk edge, and the co-orbiting vertical resonances used by Ward and Hahn are at the position of the Moon, which is outside of the disk. Once the inclination is at 10° near the Earth, tidal evolution can carry the Moon to the current distance with the orbit inclined by about the observed 5° relative to the ecliptic. Whether the Ward–Canup scheme would be as effective in a full model of the Moon's accretion has not yet been assessed. For example, it is not clear that a disk of sufficient mass could even exist after the Moon was formed or could persist long enough to increase the lunar inclination to the necessary value given the very high efficiency of accretion. Ward and Canup considered the IVR torque due to the total inner disk surface density in their analysis. This inherently assumes that the waves in both the disk's gas and magma layers damp sufficiently. If the gas waves do not damp sufficiently, then the torque could be dominated by that of the magma layer, an issue that should be addressed by future models. The Moon's resonant interactions with a Roche interior disk would generally predate its encounter with the evection resonance. An inner disk contributes to the lunar perigee precession, shifting the location of evection outward. The disk mass must be reduced to $< 0.1 M_M$ before the evection resonance moves inside $10R_E$ where it can encounter the Moon (Canup, 2004a), thereby making the Ward–Canup and Touma–Wisdom schemes mutually exclusive.

If the Moon is partially molten in the Ward–Canup scenario, the high dissipation limits the eccentricity growth in a subsequent passage through the evection resonance as the disk is dissipated and the lunar orbit continues to expand, even if trapped for a time (Touma and Wisdom, 1998). The passage through the eviction resonance is in the wrong direction for capture so the inclination will not change significantly after the Ward–Canup mechanism has brought it above 10°.

The major recent change in the giant impact paradigm has been the accommodation of the identical isotope ratios in Earth and Moon. Two successful modifications of the giant impact scenario involving an initial value of system angular momentum $\sim 2L_{EM}$ have been proposed, and there may be a way to remove the excess angular momentum. But getting from the disk of essentially identical isotope composition as the Earth's mantle to the Moon at $\sim 10R_E$ with orbit inclined 10° to the equator has many uncertain steps and generally has a deficiency of investigations. The reader can identify several processes needing more thorough investigation, and even those processes that are reasonably well understood have uncertainties that remain to be resolved. For example, how can we reconcile the initially hot Moon inferred from the giant impact origin and rapid accretion with those observations consistent with an initially cold moon? We are sure that imaginative minds will develop new ideas perhaps motivated or constrained by new data on the properties of the Moon that must be produced in a complete theory of origin.

10.17.3 Mars System

Mars' inner satellite Phobos is located at a distance of 2.77 Mars radii (R_M) from the center of Mars, which is well inside the corotation radius of $\sim 6.0 R_M$, and the outer satellite, Deimos, is just outside the corotation radius at $6.91 R_M$. The tides raised on Mars thus cause Phobos to be spiraling toward Mars and Deimos to be spiraling away. In fact, Phobos is inside the Roche radius for a density of 1.9 g cm^{-3} and would be torn apart by tidal forces if it were a fluid. It needs shear strength of only 10^5 dynes cm^{-2} to resist disruption (Yoder, 1982) – a loose rubble pile would survive (Dobrovolskis, 1982; Soter and Harris, 1977).

Determinations of the dissipation function Q (**Appendix B**) of Mars are $O(100)$ (Bills et al., 2005; Duxbury and Callahan, 1981; Shor, 1975; Sinclair, 1978) from observations of the secular acceleration of Phobos' orbital mean motion. If we choose a constant value of $Q_M = 100$ with Love number $k_M = 0.16$ (Bills et al., 2005), Deimos' orbit could have expanded by < 200 km in 4.6×10^9 years. Phobos' initial semimajor axis would have been near $5.6 R_M$ under the same assumptions. The rotation period of Mars has been essentially unaffected by the exchange of angular momentum with the satellites and would have been increased by only about 10 min due to solar tides. Deimos has essentially its initial orbit, and Phobos having started inside the corotation radius is consistent with the measured current value of $Q_M \approx 100$.

Both Phobos and Deimos are synchronously rotating; Deimos would have reached this state in $< 10^8$ years from an unlikely small initial spin period of 4 h, where a rigidity of 5×10^{11} dynes cm^{-2} and $Q = 100$ were assumed. Under the same assumptions, Phobos would have reached this state in $< 10^5$ years at its current separation from Mars and in $< 10^7$ years at its likely initial separation near the corotation radius. (See Peale (1977) for a detailed discussion of the rotation histories of all of the satellites known at that time.)

The two satellites of Mars are in nearly circular equatorial orbits and this argues for their accretion in situ from a debris disk in the equatorial plane – like the regular satellites of the major planets. However, the density of Phobos of 1.88 ± 0.07 g cm^{-3} determined from the 5-day rendezvous by Soviet spacecraft Phobos 2 with $GM_P = 7.22 \pm 0.05 \times 10^{-4}$ km^3 s^{-2} (Kolyuka et al., 1990) coupled with the volume of 5748 ± 190 km^3 from Thomas (1993) has been verified by the analysis of the flybys of the Mars Express spacecraft (1.876 ± 0.020 g cm^{-3}) (Andert et al., 2010). The density of Deimos is 1.48 ± 0.22 g cm^{-3} (Rosenblatt, 2011). Early evaluations of albedos near 5% and the reflection spectra are consistent with carbonaceous chondritic material (Pang et al.,

1978; Pollack et al., 1978). But this interpretation has been challenged by Rosenblatt (2011), who finds the spectra to be inconsistent with meteoritic material, whereas the thermal IR is consistent with silicate material. Nevertheless, these early interpretations of the spectra and the fact that the densities are much less than the mean density of Mars of 3.9 g cm^{-2} lend support to the suggestion that the satellites are composed of material that did not originate in the vicinity of Mars, although porosity is another means of reducing the density (25–45% for Phobos, 40–60% for Deimos, Rosenblatt, 2011). This interpretation inspired the consideration of intact capture (Mignard, 1981b; Pollack and Burns, 1977), but Szeto (1983) showed several seemingly insurmountable inconsistencies with the capture hypothesis independent of the impossibility of relaxing the captured satellites into the circular equatorial orbits where they are found. If the debris were indeed carbonaceous chondritic, which is not at all certain, one possible way it could have gotten into orbit about Mars would be the shattering of a planetesimal that was formed in the asteroid belt region of the nebula when it collided with a denser object already in orbit about Mars. Samples of both Phobos and Deimos would tell us if such a contrived origin were necessary. Alternatively, a giant impact on Mars could have launched a disk of material that would settle to the equatorial plane (Craddock, 2011). Craddock pointed out basins on Mars that could be the signature of such an impact. Whether an impact consistent with such basins could have produced a sufficiently extended disk to produce Phobos and Deimos at their initial locations close to the current corotation radius has not been determined. If an accretion disk is indeed the birthplace of Phobos and Deimos, many Phobos-like objects may have impacted Mars, leaving a distribution of craters with asymmetrical or elongate ejecta (Schultz and Lutz-Garihan, 1982), and we are seeing only the surviving members of the swarm. Rosenblatt and Charnoz (2012) had demonstrated the feasibility of accreting Phobos and Deimos near the corotation resonance. Any scheme to capture these satellites intact and bring them into their current orbits cannot survive close inspection of the assumptions involved.

The hypothesis that Phobos and Deimos are remnants of material in a dissipative disk, whatever the disk's origin, is not compromised by the inferred orbital history of the satellites. The initial equatorial orbits remain equatorial in spite of the chaotic, large amplitude variations in the obliquity of Mars (Laskar and Robutel, 1993; Touma and Wisdom, 1993; Ward, 1979) and despite the precession of the spin axis of Mars (Goldreich, 1965). The solid angle described by the orbit normal as the satellite orbit precesses due to Mars' oblateness is an adiabatic invariant (Goldreich, 1965), as the precession rates for the Martian satellites (periods of 2.3 and 57 years for Phobos and Deimos, respectively; Peale, 1977) are fast compared with rates of change of Mars' spin axis direction relative to inertial space (timescales $O(10^5)$ years; Folkner et al., 1997; Touma and Wisdom, 1993).

If we insist that both satellites started with nearly circular orbits, how then can we explain the current eccentricity of Phobos' orbit $e_P = 0.0151$? Both the tidal dissipation in Phobos and that in Mars due to tides raised by Phobos damp the eccentricity (see eqn [B.28]). If the orbital motion is integrated backward in time, this eccentricity grows to large values and collisions with Deimos would have been likely (Szeto, 1983; Yoder, 1982), even if there were no tidal dissipation in Phobos. Significant dissipation in Phobos reduces the timescale for a crossing orbit with Deimos to $< 10^9$ years in the past (Yoder, 1982). The current eccentricity cannot therefore be a remnant eccentricity from tidal decay beginning 4.6×10^9 years ago. Yoder (1982) had identified three commensurabilities (defined when two characteristic periods in the description of the motion are in the ratio of small whole numbers) that Phobos has passed through within the past 10^9 years that provide likely gravitational excitations of Phobos' eccentricity during its inward spiral. The commensurabilities are encountered at semimajor axis $a = 3.8$, 3.2, and $2.9R_M$, where the earliest resonance was encountered only 5×10^8 years ago. The first and third are 2:1 and 3:1 commensurabilities between the rotation of Mars and the orbital mean motion, where the resonant interaction is with Mars' axial asymmetry. At $3.2R_M$, the 2:1 commensurability is between the apparent mean motion of the Sun and the periapse motion of Phobos' orbit, where the latter's secular motion is due to the oblate figure of Mars. This resonance is like the evection resonance for the Moon. There is also a 3:2 spin-orbit resonance excitation of the eccentricity when $a = 4.6R_M$, but this excitation happened so long ago that there would be no contribution to the current eccentricity. The eccentricity would have decayed after each excitation and plausibly arrives at the current eccentricity after the series of kicks and subsequent decays (Yoder, 1982). Orbital inclination can also be excited, and even though the resonance interaction is not as strong as it is for the eccentricities, the excited inclinations do not decay. Still, the current inclinations of the orbits are consistent with the resonance passages (Yoder, 1982).

There is a condition on the dissipation in Phobos for this scenario to work. Yoder (1982) had calculated the dissipation in the satellite accounting for the tidal dissipation both due to eccentric orbit as discussed earlier and due to the forced libration of the very asymmetrical satellite. This libration was thought to have an amplitude of 3.9° (Duxbury and Callahan, 1981; Yoder, 1982) at the time and would cause twice the tidal dissipation in Phobos that would occur if Phobos were nearly axially symmetrical in the same eccentric orbit (Yoder, 1982). Recall that dissipation in Phobos and that in Mars due to tides raised by Phobos both damp the eccentricity (eqn [B.28]). There cannot be too much damping since the series of eccentricity excitations or the current eccentricity would be less than that observed. Since the dissipation in Mars can be presumed known from the measurement of Q_M, and the magnitude of the probable excitations can be reasonably estimated from the resonance passage analysis, the current value of e_P limits the contribution by Phobos. Yoder finds that $\mu_P Q_P > 3$–6×10^{12} dynes cm^{-2} or, if $Q_P \sim 100$, $\mu_P \gtrsim 3$ to 6×10^{10} dynes cm^{-2}, which is about that of ice. However, subsequent determinations of the amplitude of physical libration for Phobos are near 1° (Borderies and Yoder, 1990; Duxbury and Callahan, 1989; Jacobson, 2010; Willner et al., 2010). This reduces the dissipation by about a factor of two from what Yoder assumed, thereby relaxing the constraint on the rigidity by about a factor of 2. The properties of Phobos are not sufficiently well known to be sure that the rigidity could satisfy this new constraint, but $\mu \gtrsim 6 \times 10^{10}$ to 1.2×10^{11} is not unreasonable.

During Phobos' spiral toward Mars, it is likely that it passed through the 2:1 orbital mean motion commensurability with

Deimos. Such a passage would excite an eccentricity of about 0.002 in Deimos orbit if the eccentricity of Deimos were much smaller than this before resonance passage (Yoder, 1982). The time of this commensurability is known if the current dissipative properties of Mars have not changed substantially since the resonance encounter. This places a lower bound on the dissipation in Deimos if the current eccentricity (0.0002) is the tidal remnant from an initial value of 0.002 excited by the resonance passage. Yoder (1982) found $\mu_D Q_D(1-\alpha_D)^2/\alpha_D^2 \lesssim 10^{10}$ dynes cm^{-2}, where $\alpha_D = 3(B-A)/C$ with $A<B<C$ being the principal moments of inertia of Deimos. This limit may be unreasonably low, but the dissipation in Deimos may be increased if the forced libration is nearly resonant with the free libration. The enhanced amplitude of libration would lead to higher dissipation and relax the constraint on $\mu_D Q_D$. The free libration period could be better constrained by an estimate of α_D from a more accurate determination of Deimos' shape along with an accurate measure of its physical libration amplitude. With semi-axes of 7.5, 6.2, and 5.4 km for Deimos (Duxbury and Callahan, 1989) and a mean density of 1.5 g cm^{-3} (Rosenblatt, 2011), $\alpha_D = 0.354$ and $\mu_D Q_D \lesssim 3.3 \times 10^{10}$ dynes cm^{-2}. A homogeneous Deimos modeled as a triaxial ellipsoid would have a free libration period about 1.3 times the orbital period, which is not sufficiently close to resonance to enhance the forced libration significantly, but the fact that Deimos is a porous rubble pile may enhance the dissipation above that for a less rigid solid body.

In any case, Yoder's hypothesis that the satellite orbits have always been regular and current properties of the system then attributed to the effects of resonance passages by Phobos is reasonably well supported. This is consistent with our presumed origin from a dissipative disk of small particles, although the details of the origin of the disk have yet to be constrained.

10.17.4 Jupiter System

The Galilean satellites of the Jupiter system and those small satellites inside Io's orbit are nearly coplanar with Jupiter's equator implying their formation in a dissipative disk (**Appendix A**). Additional constraints on the formation of the Galilean satellites of Jupiter have been established by theoretical analysis and the Galileo spacecraft observations. The mean densities of Io, Europa, Ganymede, and Callisto are, respectively, 3.53, 3.01, 1.94, and 1.83 g cm^{-3}. Ganymede and Callisto are about 40% ice and 60% rock by mass (Sohl et al., 2002). Europa has a layer of water and ice 80–170 km thick (Anderson et al., 1998b). CO_2 has been detected on the surfaces of Ganymede and Callisto (Hibbitts et al., 2000, 2002, 2003; McCord et al., 1998) and in the atmosphere of Callisto (Carlson, 1999). The discovery of an intrinsic dipole magnetic field on Ganymede and the low amplitude of higher-order multipoles in the field implies generation of the field deep within the satellite probably by dynamo action in a molten metallic core (Kivelson et al., 1997, 1999, 2002). Induced fields in Europa and Callisto and possibly Ganymede from the variation in the external field due to Jupiter's rotation are consistent with induced fields expected from a conducting liquid water layer at rather shallow depths (Kivelson et al., 2002). It has been speculated that NH_3 is sufficiently abundant on Ganymede and Callisto to lower the solidus temperature of an NH_3–H_2O mixture so that such a liquid layer could be maintained over the age of the solar system (Mousis et al., 2002; Spohn and Schubert, 2003). However, there has been no NH_3 nor N_2 detected on any of the Galilean satellites, but McKinnon (2006) argued that halides in solution may depress the solidus temperature sufficiently to preserve the liquid layers.

The large amount of relatively volatile material in Ganymede and Callisto implies that the disk temperature remained low enough during the accretion process that icy particles could persist. Callisto may not be fully differentiated (Anderson et al., 1998a, 2001; Nagel et al., 2004), and an accretion timescale $\geq 6 \times 10^5$ years is required to keep ice in its interior from melting (Barr and Canup, 2008). The change in composition from rocky to icy from Io to Callisto is consistent with an expected radial temperature gradient in the disk at the time of satellite accretion, but note the existence of the icy satellite Amalthea inside Io's orbit. It is possible that Europa lost initial volatiles due to tidal heating (**Appendix B**). Given its current tidal dissipation rate, a very efficient conversion of tidal heating into vaporization of ice is required over the satellite's entire lifetime in order for Europa to have started as ice-rich as Ganymede and Callisto (e.g., Canup and Ward, 2009). If Europa's ice depletion is primordial, then Io's likely is as well, since by virtue of being interior to Europa, it may have formed in a higher-temperature environment.

The theories for the formation of the Galilean satellites are guided by numerical hydrodynamic calculations – first in two dimensions (2-D) (e.g., Lubow et al., 1999) and later in 3-D (e.g., Bate et al., 2003; d'Angelo et al., 2003; Machida et al., 2008; Tanigawa et al., 2012), where the flows near the Hill sphere are strikingly different in the two types of calculations. It is clear that 3-D calculations are required to constrain the conditions of satellite formation in planetary accretion disks, but the incompleteness of these calculations (e.g., there are no numerical calculations that follow the development of the disk from the collapse phase of Jupiter's atmosphere to the satellite formation stage) has led to a variety of models involving speculation of the disk properties and the dominant processes therein.

The formation of the regular Jovian satellites must necessarily occur late in Jupiter's accretion to allow the hot inflated protoplanet to cool and contract to a radius less than that of the closer satellites. This can plausibly occur by the time the planet is accreting the last few tens of percent of its mass, depending primarily on the planet's opacity (Ward and Canup, 2010). Minimum mass disk models (called minimum mass sub nebula (MMSN) models) for the formation of the satellites require sufficient mass of condensable material to form the satellites augmented with gas to reach solar composition in the initial disk (e.g., Lunine and Stevenson, 1982), where no mass is added to the disk during the satellite accretion process. Mosqueira and Estrada (2003a,b) modified the MMSN model by enhancing the fraction of solid material in the disk (solids-enhanced minimum mass), so there is less gas than in the traditional MMSN models. Various theoretical considerations of disk processes and histories (**Appendix B**) will lead us to favor models (with some caveats) where the current satellites form during the very last stages of Jupiter's

accretion when the accretion disk was 'starved' of gas input (Alibert et al., 2005a,b; Canup and Ward, 2001, 2002, 2006, 2009; Mousis and Alibert, 2006; Ogihara and Ida, 2012; Sasaki et al., 2010; Stevenson, 2001a,b; Ward and Canup, 2010). In these latter models, considerably more condensable mass passes through the disk than is contained in the current satellites, and these objects form in a disk with several orders of magnitude less gas than in an MMSN.

The shocks near the L_1 and L_2 Lagrange points, after a gap is formed in the solar nebula, which are found in the 2-D calculations of Bryden et al. (1999), Kley (1999), and Lubow et al. (1999), are modified considerably in the higher-resolution 3-D calculations of Machida et al. (2008). For a 1 Jupiter mass protoplanet, the shocks are displaced about 45° from the Sun–Jupiter line and stand off from the Hill sphere by about 20% of R_H. A reservation about this comparison is that the calculations correspond to a full density solar nebula and are local. The local nature of the Machida et al. calculations means the creation of a deep gap in the disk as Jupiter's mass grows cannot be treated (Machida et al., 2010). In the final stages of accretion, the circumplanetary disk is more likely to behave like a standard accretion disk rather than being subject to the shock-driven accretion found in the 2-D calculations (Bate et al., 2003). In both 2-D and 3-D calculations, a prograde accretion disk is formed, which is the basis for the formation of a coplanar set of satellites in nearly circular orbits.

In spite of Jupiter having likely opened a gap (**Appendix A**) in the solar nebula when its mass exceeded approximately 1/3 of its current value (e.g., Bryden et al., 1999; Ward and Hahn, 2000), 3-D numerical simulations (e.g., Kley et al., 2001) find continuing accretion onto Jupiter at a rate of 10^{-4}–$10^{-2}\,M_E$ year^{-1}, depending on the assumed nebular viscosity, surface mass density, and scale height. Bate et al. (2003), in a 3-D hydrodynamic simulation, found an accretion rate close to $10^{-4}\,M_J$ year^{-1} ($3\times10^{-2}\,M_E$ year^{-1}) as Jupiter approaches its final mass, M_J. The latter authors find that accretion only stops when the planet at 5.2 AU reaches about $10M_J$, with an undepleted solar nebula. The accretion of Jupiter was therefore most likely protracted, with the rate of accretion tailing off as the solar nebula is depleted. A striking conclusion from the 3-D calculations (Bate et al., 2003; Machida et al., 2008; Tanigawa et al., 2012) is that relatively little material from the midplane of the solar nebula is accreted onto the disk. Most of the material rains down on the disk from above and below the midplane. This phenomenon is incorporated into starved disk models discussed in the succeeding text, but first, we describe a recently developed solids-enhanced minimum mass (SEMM) model (Estrada et al., 2009; Mosqueira and Estrada, 2003a,b) and a gas-free accumulation of the satellites from planetesimals captured from the heliocentric swarm (Mosqueira and Estrada, 2006).

In the following discussions of gaseous disk models, we will use the turbulent viscosity parameter α, defined by $\nu_t=\alpha c_s$, where ν_t is a characteristic velocity of a turbulent eddy and c_s is the speed of sound in the gas (**Section A.2**). The value of α is very uncertain, but typical values are chosen near 10^{-3}. Sources of turbulence in the circumplanetary disks include the influx of material onto the disk (e.g., Cassen and Moosman, 1981) and/or turbulence generated by the Magnetohydrodynamic (MHD) instability (Turner et al., 2014). The latter source requires a certain degree of ionization of the disk material that may be difficult to attain in the environments of most planetary accretion disks. See Turner et al. (2014) for a thorough discussion of the conditions where MHD turbulence may be effective in generating a viscous evolution of a disk and Lubow and Martin (2013) who consider the possibility of a layered disk in which the disk's midplane is nonturbulent, while the outer layers are MHD-active.

10.17.4.1 SEMM Model

We first outline the motivation for the initial conditions in the Mosqueira–Estrada (2003a,b) model. The angular momentum of a planetesimal (or a small volume of gas) relative to the Jupiter embryo is $\mathbf{L}=m\mathbf{r}\times\mathbf{v}$, where m, $\mathbf{r}$, and $\mathbf{v}$ are mass, position, and velocity relative to a frame centered on Jupiter but not rotating. If we write $\mathbf{v}=\mathbf{v}_{rel}+\mathbf{n}_J\times\mathbf{r}$, where $\mathbf{v}_{rel}$ is the velocity of the particle relative to a frame centered on Jupiter and rotating with the mean orbital angular velocity $\mathbf{n}_J=\sqrt{GM_\odot/a_J^3}\,\boldsymbol{e}_z$ with G, $M_\odot$, a_J, $\boldsymbol{e}_z$ being the gravitational constant, solar mass, semimajor axis of Jupiter's orbit, and normal to the orbital plane, respectively, $\mathbf{L}=m\mathbf{r}\times(\mathbf{v}_{rel}+\mathbf{n}_J\times\mathbf{r})=m\mathbf{r}\times\mathbf{v}_{rel}+mr^2\mathbf{n}_J$, where the second form is for the assumption that $\mathbf{r}$ is in the plane of the disk and thereby perpendicular to $\mathbf{n}_J$ along the orbit normal. The specific angular momentum is

$$\ell=\mathbf{r}\times\mathbf{v}_{rel}+r^2\mathbf{n}_J \qquad [8]$$

where the first term is the specific angular momentum relative to the rotating system and the second is the correction for inertial space representation.

Before the creation of a gap in the heliocentric nebula, ($M\ll M_J$), it is assumed that all of the flux encountering the Hill sphere of radius $R_H=(M/3M_\odot)^{1/3}a$ is captured by the proto-Jupiter. Lissauer and Kary (1991) found for the sum of all the specific angular momenta

$$\ell=n_J\frac{\int_0^{R_H}-1.5x^3\,dx}{\int_0^{R_H}x\,dx}+n_JR_H^2=\frac{1}{4}n_JR_H^2 \qquad [9]$$

where x is the difference in the semimajor axes of the particle and Jupiter, where we have dropped the vector notation, since it is assumed that all contributions to the specific angular momentum of accreted material are perpendicular to the disk. With conservation of angular momentum, the centrifugal radius is defined by $r_c=\ell^2/GM=R_H/48\approx15R_J$ for the value of ℓ given in eqn [9] and for R_H determined for $M=M_J$, where R_J is Jupiter's current radius. The centrifugal radius r_c is interpreted as the disk size. For $M\ll M_J$ appropriate for eqn [9], the disk before gap opening is quite compact.

After the gap in the heliocentric disk becomes large compared to R_H, material enters the Hill sphere near the Lagrange points in the 2-D calculations of Artymowicz and Lubow (1996) and Lubow et al. (1999). In the 3-D calculations of Machida et al. (2008), material enters the Hill sphere, after passing through the shocks, in the first and third quadrants projected onto the $z=0$ plane with the x-axis pointing away from the Sun. Nearly all of the 3-D calculations show material falling vertically onto the disk (d'Angelo et al., 2003; Machida,

2009; Machida et al., 2008, 2010; Tanigawa et al., 2012). After gap opening, Mosqueira and Estrada (Estrada et al., 2009) assumed that gas enters the Hill sphere at very small relative velocity. The first term in eqn [9] thereby vanishes and the specific angular momentum becomes $R_H^2 n_J$ and $r_c = R_H/3 \sim 250 R_J$.

Based on this, Mosqueira and Estrada (2003a,b) proposed a two-component disk: a massive disk for ($r \lesssim 15 R_J$) that persists before gap opening surrounded by an extensive outer disk ($r \sim 150 R_J$) of much lower surface mass density, emplaced after a large gap has formed. The sparseness of the outer disk follows from the much reduced gas flow into the Hill sphere after the gap approaches its maximum extent, and the large radius of the disk follows from the higher specific angular momentum of the material entering the Hill sphere at the Lagrange points as discussed earlier. The gas influx rapidly goes to zero due to gap opening, but solid material continues to be added to the disk through ablation of fragmented planetesimals passing through the disk (e.g., Mosqueira et al., 2010) or collisions of such fragments with existing solid satellitesimals already in the disk (e.g., Estrada and Mosqueira, 2006, see succeeding text).

The gas flow is assumed to wane on the same timescale as the gap opening timescale of $\sim 10^3$ years for a weakly turbulent heliocentric disk (Bryden et al., 1999, 2000; Morbidelli and Crida, 2007). The cessation of the gas flow onto the disk means the presumed chief source of turbulence in the disk is eliminated and the disk viscosity and temperature plummet. Note that the gas flow ends when $M \approx M_J$, which is contrary to the conclusion of Bate et al. (2003), where accretion stops only when $M \approx 10 M_J$, although the latter is for an undepleted solar nebula.

The postaccretion low disk viscosity and low temperature are involved in various schemes for avoiding the short timescales for the important processes in the disk discussed in **Appendix A**. The system is closed, since continuing accretion is omitted, where the minimum mass tag implies there is only sufficient solid material to make the satellites, but the gas content of the disk is down from a solar composition by a factor of 10 (Estrada et al., 2009). A $1/r$ temperature distribution is assumed with $T = 250$ K at Ganymede's distance in order to accommodate the stability of icy particles at greater distances. A feature of the model is a surface density distribution ($\sigma_g \propto 1/r$) with sufficient solids to make Io, Europa, and Ganymede inside $\sim 15 R_J$, leading to an optically thick, massive inner disk. Mosqueira and Estrada set the value of the gas surface density to $\sigma_g \sim 3 \times 10^4$ to 3×10^5 g cm^{-2}, with the lower end of the range indicating a disk enhanced in solids compared to the standard MMSN disk. But the surface density falls drastically beyond about $15 R_J$ to a value determined by spreading Callisto's mass in the outer disk out to $r \sim 150 R_J$. With an opacity due only to the gas (i.e., without dust grains), an optically thin outer disk results from which it is argued that Callisto is accreted slowly ($\sigma_g \sim 1/r$ for $r > 26 R_J$). The $1/r$ temperature dependence in the inner disk goes to a constant temperature in the outer disk of 130 K, like that of the solar nebula. A controversial assumption in the model is that the sharp gradient in the surface density between the inner and outer disks is maintained throughout the accretion timescale of Callisto, which they estimate to be $O(10^5$–10^6 years).

The steep surface density gradient between the inner and outer disks is maintained by assuming that turbulence has been nearly completely damped leading to $\alpha \sim 10^{-5}$–10^{-6} in both the inner and outer disks. Such low values of α might not be definitely ruled out, since there is no continuing accretion to stir the disk. Small but decoupled objects are rapidly lost to Jupiter through gas drag. But the accumulation of dust and rubble into larger objects is more rapid than the decay of these objects by gas drag. The small particles are assumed entrained in the gas and encounter larger particles that are at Kepler velocities. The growth timescale is $M/\dot{M} = 4\rho_p R(3\rho_s \Delta v)$, where Δv is the difference between the Kepler and gas velocities. Sweepup time is smaller than gas drag time for $r < 38 R_J$ independent of particle size or distance range where $T \propto 1/r$, such that satellite formation is favored. Then, the timescale for accretion is determined by the rate at which gas drag can bring say 100 m or larger objects to the protosatellite. Accretion times for Ganymede under these assumptions are $O(10^3$–$10^4)$ years. Since they have little or no ice, Io and Europa can accrete more rapidly than Ganymede.

The short migration timescales indicated in **Appendix A** are frustrated by the assumption that the inner three Galilean satellites open gaps in the disk and change from type I to type II migration. The latter is on the viscous timescale, which exceeds the supposed timescale for elimination of the gas disk by photoevaporation (Hollenbach et al., 2000; Shu et al., 1993), because the turbulent viscosity characterized by $\alpha = O(10^{-6})$ is so small. The steep gradient in the surface mass density between about 20 and $26 R_J$ causes type I drift to reverse sign in this region so that a satellite would spiral out instead of in. Callisto is hypothesized to have ended up where the type I drift vanishes at the value of the density gradient where torques from the inner Lindblad resonances balance those from the outer ones. The steep gradient in σ_g between the inner and outer disks must be maintained not only through the accretion phase but also throughout the dissipation of the disk. Alternatively, Callisto could open a gap, shift from type I to type II drift and thereby end up where it is because of the low viscosity assumed. But why did it wait until $r = 26 R_J$ to open the gap?

The SEMM model remains at this point a conceptual framework for satellite formation; there have been no explicit, time-dependent models demonstrating either the formation of the assumed SEMM initial conditions or that the proposed SEMM disk would ultimately accrete into a system of four Galilean-like satellites. Many of the subjective criticisms of the Mosqueira–Estrada model, such as the very low values of α assumed from the outset to preserve the steep surface density gradient between the inner and outer disks, are muted by uncertainties about which processes are ongoing and on the values of various parameters therein. However, many aspects of the model are ad hoc, such as the $1/r$ temperature distribution normalized to allow ice to condense just beyond Ganymede's orbit, the low viscosity (and maintenance of the steep surface density gradient between the inner and outer disks), and the assumption of the existence of a minimum mass, closed disk in the first place. The latter assumption ignores the history of the disk during the long period of waning gas accretion. How could the massive inner disk resist viscous

spreading during this period when accretion still generated a reasonable turbulent viscosity?

The SEMM model, as well as the model of Sasaki et al. (2010) discussed in the succeeding text, is not consistent with recent results from increasingly sophisticated numerical simulations of planetary accretion (e.g., Bate et al., 2003; Machida et al., 2008), which show the protracted nature and persistence of the accretion process as the solar nebula is dissipated. These latter results confirm intuitive conclusions about the accretion histories of the giant planets. The neglect of this history and the added criticisms in the preceding text mean that the SEMM and MMSN models in general are not likely to represent the actual sequence of events that led to the formation of the regular natural satellites around the giant planets. In spite of these criticisms, we note that the Mosqueira–Estrada papers contain discussions of many, if not most, of the processes that occur in disks that effect the satellite development. These processes should be studied in the development of a comprehensive model of origin of the regular satellites.

10.17.4.2 Planetesimal Capture Model

In a rather detailed and careful analysis, Estrada and Mosquiera (2006) reexamined the hypothesis that collisions of planetesimals within Jupiter's Hill sphere in a gas-free environment led to a buildup of captured orbiting material that settles to an equatorial disk through mutual collisions where they could have accumulated into the satellites (e.g., Safronov et al., 1986). Constraints are placed on numerous parameters and assumptions that could allow such a scenario to produce the observed masses and angular momentum distributions, but the authors note that whether or not the parameter values and assumptions are likely to obtain is very uncertain. Perhaps the greatest value of this paper is to show how difficult it is to make a gas-free planetesimal capture scenario a viable alternative for satellite formation. Major caveats are the difficulty in producing the composition gradient in such a scenario (unless later Roche lobe interlopers or tidal dissipation preferentially removes volatiles from the inner satellites), keeping a size distribution of planetesimals that favors collisional capture, and maintaining the supply of planetesimals within Jupiter's feeding zone over the 10^5–10^6-year timescale of the satellite accumulation. It is not clear that a sufficient gas density in the solar nebula to allow a gas drag replenishment of planetesimals in Jupiter's feeding zone would not also keep a gas accretion disk replenished by streaming across the gap. A problem with planetesimals later colliding with the accreted satellites to preferentially remove volatiles from the inner satellites is attaining a sufficient flux of such planetesimals from the depleted zone around Jupiter.

Another problem that is not addressed is the reduction of the optical depth in the circumplanetary disk as the satellitesimals accumulate into larger objects on short timescales, thereby reducing the collisional capture probability of additional mass. In addition, the scenario is not placed in the context of the formation of Jupiter that yields a plausible route to the assumed gas-free initial conditions where Jupiter has no satellites. Presumably, Jupiter has already gone through the last stages of gas accretion through streams of material across a gap in the solar nebula, and we shall see later that it should have acquired a stable satellite system already during that process. With the caveats pointed out by Estrada and Mosqueira and the additional ones we have mentioned here, it seems unlikely that the satellites formed solely from collisionally captured planetesimals in a nearly gas-free environment.

10.17.4.3 Starved Accretion Disk Model

Accumulation of the satellites of Jupiter in an actively supplied, gas-starved accretion disk instead of a MMSN was proposed by Canup and Ward (2001) and separately by Stevenson (2001a) at the Jupiter conference in Boulder, Colorado, in June 2001 (http://lasp.colorado.edu/jupiter/index.html). In this construct, the planet contracts to a size smaller than the satellite region before the nebula dissipates, with subsequent inflow producing a circumplanetary accretion disk (e.g., Ward and Canup, 2010). The satellites form in the disk as the inflow is ongoing, rather than after the inflow has stopped (as in the MMSN and SEMM models). As the planet nears its final mass and the inflow wanes, the disk becomes increasingly gas-starved. Motivations for the starved disk model include a more consistent treatment of the conditions of satellite formation based on planet formation models, the necessity of disk temperatures low enough to allow water ice to condense, and a sufficiently slow accretion of Callisto to allow possibly incomplete differentiation (e.g., Anderson et al., 2001; Kuskov and Kronrod, 2005; McKinnon, 1997; Schubert et al., 2004). Stevenson (2001b) summarized remarks at the conference, whereas Canup and Ward (2002) developed a detailed steady-state model based on their conference paper. Their model was extended by Alibert et al. (2005a,b) who accounted for the time variation of disk properties as accretion wanes. Major features in the Alibert et al. approach are that material from the solar nebula is deposited at the outer edge of the disk with a mass flux through the disk that is in quasi-steady state and that the last part of the satellite formation occurs in a closed disk after accretion has ended. Canup and Ward (2002) had the mass raining down onto the disk consistent with a major finding of the 3-D calculations of Bate et al. (2003), Machida et al. (2008), and Tanigawa et al. (2012) showing the vertical deposition. Canup and Ward (2006) extended their model to include an exponential decrease in the accretion rate with time and a direct model of satellite growth, in which many of the accreted satellites are lost to Jupiter from migration and only the last set formed near the end of accretion remain. Ward and Canup (2010) considered the formation of a disk around a growing gas giant planet, including the energy and angular momentum budget of the inflowing material and the coupled thermodynamic evolution of the planet itself. They found that by the time the planet accretes the final ~10% of its mass, an accretion disk with low gas surface densities and properties broadly similar to those considered in Canup and Ward (2002, 2006) has formed. Sasaki et al. (2010) and Ogihara and Ida (2012) considered two variations on the Canup–Ward model: an inflow that ends abruptly rather than exponentially due to gap opening (and in this these two works share a common element with the SEMM model) and a disk with a magnetically sustained inner cavity. It is the Canup–Ward (2002, 2006, 2009) model that we now describe.

Ward and Canup (2010) argued that the vertical deposition of material onto the disk and planet falls in the range $r_c/4 < r < 9r_c/4$, where the upper bound is close to the $30R_J$ used in their 2002 model. For accretion during gas inflow, the steady-state disk with uniform deposition inside a radius of $30R_J$ treated in 2002 is generalized in 2006 to have a time-dependent influx density of

$$F_{in} \propto \left(\frac{1}{r}\right)^{\gamma_{in}} \exp\left(-\frac{t}{\tau_{in}}\right) \quad [10]$$

where F_{in} is the vertical flux density onto the disk's upper and lower surfaces, r is the radial distance from the center of Jupiter, τ_{in} is the time constant for exponential decay of the influx rate, and γ_{in} is an adjustable constant parameterizing the radial variation in the flux density over the disk inside the outer bounding radius. The radius r_c is a free parameter constrained by the specific angular momentum of inflowing material. A range of 16–46 planetary radii for Jupiter, Saturn, and Uranus is consistent with Lissauer's (1995) 3-D calculation (with 16 being the value for Jupiter in particular). The disk outer boundary is set to $r_d \approx 150R_J$. If the viscous timescale $\tau_\nu \approx r^2/\nu$, with $\nu = \alpha c_s H$ (c_s = midplane sound speed and H = gas scale height, **Appendix A**), is much smaller than τ_{in}, the gas surface density σ_g maintains a quasi-steady state like that assumed in Canup and Ward (2002). For $\tau_{in} \sim 10^6$ years, this condition is valid for $\alpha \geq 10^{-4}$.

Solids smaller than ~1 m are entrained in the gas and, for a planet that has not opened a gap, will be delivered with it to the disk from heliocentric orbit. The delivery of small particles to the disk is more complex once the planet opens a gap. At the outer gap edge, the gas disk's rotation is super-Keplerian because the local gas surface density increases with orbital radius, thereby reversing the radial pressure gradient. The gas drag that caused the inward migration of small particles is reversed, causing them to stall and build up at the outer gap edge (Paardekooper and Mellema, 2006a,b). However, this buildup cannot proceed indefinitely, because once the surface density of small particles becomes comparable to that of the local gas, the back reaction of the particles on the gas causes its motion to become more Keplerian and particles begin to flow into the gap (Ward, 2009). A simple model accounting for this effect suggests that particles mm-sized and smaller could enter the gap and be delivered to the disk (Ward, 2009), although this issue merits further investigation for larger particles.

It has traditionally been thought that most solids in the heliocentric disk would be in large planetesimals by the time of Jupiter's formation, although these would probably be diminished in size through fragmentary collisions after Jupiter had gained most of its mass. However, recent works argue that a large fraction of the disk's solid mass was contained in millimeter- to centimeter-sized 'pebbles,' based both on observations of extrasolar disks (e.g., Testi et al., 2003; Wilner et al., 2005) and on dynamic models (e.g., Lambrechts and Johansen, 2012). The latter found that pebble accretion can lead to formation of the gas giant cores in a time shorter than the nebular lifetime. Such conditions would also imply a substantial population of pebble- and grain-sized particles that could be delivered to the circumplanetary disk by the gas inflow.

It may also be the case that an alternative delivery of solids to the disk may be viable. Once in the disk, the solids are assumed to accrete rapidly enough to decouple from the gas before they can spiral into Jupiter from gas drag. In an inflow-supplied disk, the accretion timescale for a satellite of mass m_s is $\tau_{acc} \approx f m_s/(2\pi r \Delta r F_{in})$, where $f \approx 100$ is the gas to solid ratio for the influx and Δr is the width of the annulus over which the satellite accumulates material. With $\Delta r/r \approx 2e$, where $e \approx (H/r)(m_s/4\pi r H \sigma_g)^{1/3}$ is the maximum eccentricity obtained by m_s from a balance between eccentricity damping by density waves and excitation by scattering from similarly sized objects (Ward, 1993), the accretion timescale for a uniform inflow varies with orbital radius as $\tau_{acc} \propto 1/r^{4/5}$ (Canup and Ward, 2006). The inverse dependence on r is because accretion is being regulated by the rate of the inflow, rather than by the orbital frequency as is the case in disks with a fixed total mass. For a uniform inflow per area, mass is added more quickly to an outer satellite's feeding zone due to its much larger surface area, so that objects in the outer disk grow more quickly than those in the inner disk.

Satellite growth continues, while each satellite spirals toward Jupiter from type I drift (**Appendix A**). The rate of type I drift is proportional to the satellite mass and to the gas surface density in the disk (**Appendix A**), and the satellites tend to accelerate as they drift toward Jupiter due to their accretion of some additional inner disk material. The timescales for type I drift (**Appendix A**) are generally shorter than τ_{in} in eqn [10], so that many satellites are lost to Jupiter as the influx winds down. But as the inflow wanes, the gas surface density drops and the type I timescale lengthens; the last surviving satellites are then those for which $\tau_1 \sim \tau_{in}$. Each satellite reaches a critical mass m_{crit} where the timescale τ_{acc} for further growth is comparable to its type I orbital decay timescale after which further growth is limited before the satellite is lost to Jupiter.

The model includes disk heating and cooling, where the heating comes from the gravitational energy lost by the incoming material, viscous dissipation within the disk, and even radiation from a still warm Jupiter with only radiative cooling to keep the disk temperature below the ice point at least in the outer regions. Disk temperature distributions and surface densities are constrained by assumptions about the disk opacities and the viscosity parameter α defined in **Appendix A**. The timescale for type I drift depends on the scale height, which depends on the temperature of the disk, and that temperature distribution depends on the opacity of the solids–gas mixture in the disk, which itself is a function of temperature (Pollack et al., 1994; **Appendix A**). Higher opacities lead to slower required inflow rates for similar temperatures. Opacity is then important in determining both the disk structure and the manner in which the migration rate changes with r. Unfortunately, the opacity and its dependence on radial distance from Jupiter are highly uncertain, since it depends on the amount, composition, and size distribution of the dust in the nebula. Canup and Ward (2002, 2006) used a constant opacity in their model, and it would be of interest to determine the effect of other assumptions about the opacity distribution and the general thermal properties of the disk on their conclusions.

For the case where $\gamma_{in} = 0$ in eqn [10] (uniform deposition inside r_c) Canup and Ward (2006) find in their eqn [2],

$$\frac{m_{\rm crit}}{M_{\rm J}} \approx 5.4\left(\frac{\pi}{C_{\rm a}}\right)^{5/9}\left(\frac{H}{r}\right)^{26/9}\left(\frac{r}{r_{\rm c}}\right)^{10/9}\left(\frac{\alpha}{f}\right)^{2/3}(\Omega_{\rm K}\tau_{\rm G}f)^{1/9}$$
$$\approx 5.6\times10^{-5}\chi\left(\frac{3.5}{C_{\rm a}}\right)^{5/9}\left(\frac{H/r}{0.1}\right)^{26/9}\left(\frac{r/r_{\rm c}}{0.5}\right)^{10/9}\left(\frac{\alpha/f}{3\times10^{-5}}\right)^{2/3} \quad [11]$$

where $\chi=[(1\ \text{week}/\{2\pi/\Omega_{\rm K}\})(\tau_{\rm G}/10^7\ \text{years})]^{1/9}$ and $\tau_{\rm G}=M_{\rm J}/(dM_{\rm J}/dt)$ is the timescale for the growth of $M_{\rm J}$. Because the ratio $m_{\rm crit}/M_{\rm J}$ depends so weakly on the inflow rate ($\tau_{\rm G}^{1/9}$), a similar maximum satellite mass results for a wide range of inflow rates. Next, the dependence of $m_{\rm crit}/M_{\rm J}$ on α/f for a given inflow rate can be understood from the following. A larger α leads to a lower surface mass density and hence slower migration, which allows the satellite mass to grow more before it is lost, and a smaller f means more solids are delivered also resulting in an increased satellite mass.

Canup and Ward (2002, 2006) used a type I timescale appropriate for optically thin, isothermal disks (Tanaka et al., 2002), with $C_{\rm a}=2.7+1.1\zeta$, where ζ is the exponent for the radial surface mass distribution as in $\sigma_{\rm g}\propto 1/r^{\zeta}$. $C_{\rm a}=3.5$ for a disk with $\zeta=3/4$ (Canup and Ward, 2002). It is possible that even a gas-starved disk can be optically thick, depending on the abundance and size of dust in the disk. Recent works have shown that type I migration is altered in optically thick, adiabatic disks due to opposing effects of corotation torques (e.g., Kley and Crida, 2008; Paardekooper and Mellema, 2006a,b). Paardekooper et al. (2010) provided a revised $C_{\rm a}$ accounting for such effects, where

$$C_{\rm a}=-\frac{2}{\gamma_{\rm a}}\left[-2.5-1.7\beta+0.1\zeta+1.1\left(\frac{3}{2}-\zeta\right)+7.9\frac{\beta-(\gamma_{\rm a}-1)\zeta}{\gamma_{\rm a}}\right]$$

where $\gamma_{\rm a}$ is the adiabatic index (1.4 for diatomic hydrogen), β is the exponent in the disk temperature profile $T_{\rm d}\propto 1/r^{\beta}$, and $C_{\rm a}>0$ (<0) is inward (outward) migration. For values appropriate for an optically thick inflow-supplied disk, $C_{\rm a}=0.5$, indicating inward type I migration that is about seven times slower than that estimated by Tanaka et al. (2002) for optically thin, isothermal disks. Because the satellite mass is set by a balance between satellite growth and type I loss, a reduction in $C_{\rm a}$ to 0.5 in eqn [11] would produce a comparable critical satellite mass for an α/f ratio that is a factor of 5 smaller, reducing, for example, the needed level of viscosity to achieve a desired satellite mass.

In these analytic approximations, each satellite grows to mass $m_{\rm crit}(r)$ within an annulus of width Δr, after which it is rapidly lost from type I migration. But in the time to migrate into Jupiter, another $m_{\rm crit}$ accumulates. So, the total mass $M_{\rm T}$ in satellites at any time is given by eqn [4] of Canup and Ward (2006),

$$\frac{M_{\rm T}}{M_{\rm J}}=\int_{R_{\rm J}}^{r_{\rm c}}\frac{m_{\rm crit}/M_{\rm J}}{\Delta r}dr$$
$$\approx 2.1\left(\frac{\pi}{C_{\rm a}}\right)^{4/9}\left(\frac{H}{r}\right)^{10/9}\left(\frac{\alpha}{f}\right)^{1/3}\frac{1}{(\Omega_{\rm K}\tau_{\rm G}f)^{1/9}} \quad [12]$$
$$\approx 2.5\times10^{-4}\frac{1}{\chi}\left(\frac{3.5}{C_{\rm a}}\right)^{4/9}\left(\frac{H/r}{0.1}\right)^{10/9}\left(\frac{\alpha/f}{3\times10^{-5}}\right)^{1/3}$$

where it is assumed that H/r and f are approximately constant across the disk and that $R_{\rm J}\ll r_{\rm c}$. An important characteristic of eqn [12] is that $M_{\rm T}/M_{\rm J}$ is comparable to the observed satellite–planet mass ratios for Jupiter with plausible choices for the parameters, and by replacing $M_{\rm J}$ by $M_{\rm S}$ or $M_{\rm U}$, the same applies to Saturn and Uranus. Notably, $M_{\rm T}/M_{\rm J}$ is insensitive to the inflow rate through χ, lacks a dependence on $r_{\rm c}$, and depends only weakly on α/f.

Perhaps the most notable result of this model is the prediction that all the major planets should have values of $M_{\rm T}/M_{\rm P}\approx10^{-4}$, which is what is observed. The value of $M_{\rm T}/M_{\rm P}$ is determined primarily by the value of α/f, and the weak dependence of the mass fraction on this ratio ($(\alpha/f)^{1/3}$) means that systems comparable to those of Jupiter, Saturn, and Uranus result for $10^{-6}<\alpha/f<5\times10^{-4}$. Neptune is excluded because the capture of the large satellite Triton disrupted the original system. The application of the theory to Uranus requires additional processes that lead to the large obliquity after the satellite system was formed. An impact that tilted Uranus would disrupt the Canup and Ward-formed satellite system leading to mutual collisions among the satellites and reassembly in the new equatorial plane, or if an adiabatic process can be found to tilt Uranus, the satellite system would remain equatorial during the process (see **Section 10.17.6**). Both α and f are most likely to vary both in time and from planet to planet given the difference in the planet compositions and the likely decrease in f from solar composition as the solar nebula is evaporated. Whether or not the given influx of material can cause enough turbulence for values of α/f within this range has not been ascertained. Magnetorotational instability is another potential source of turbulence, but its effectiveness is greatly diminished by the likely presence of dust grains and likely insufficient ionization (e.g., Fujii et al., 2011).

To check these conclusions, satellite formation is simulated in Canup and Ward (2006) by the N-body SyMBA code (Duncan et al., 1998) based on the algebraic mapping integration scheme of Wisdom and Holman (1991), modified to include the interaction of the masses with the gas disk and ongoing mass inflow (Canup and Ward, 2006). The satellites interact gravitationally with Jupiter and with each other. The background disk is treated analytically and interactions of the satellites with it are treated as small perturbations of the orbits. The radial surface density profile is approximated by $\sigma_{\rm g}(r)=\sigma_0(R_{\rm J}/r)^{\gamma_{\rm g}}$, and $H/r=c_{\rm s}/r\Omega_{\rm K}=h_0(r/R_{\rm J})^{\gamma_{\rm c}}$ gives the radial variation of the scale height. The scale height depends on the radial temperature distribution determined by the opacity. Values of $\gamma_{\rm g}=0.75$, $\gamma_{\rm c}=0.13$, and $h_0=0.07$ are estimated from the steady-state disk model of Canup and Ward (2002) with a constant opacity (see **Appendix A**). Satellites experience a drag force because the gas is partially supported by the radial pressure gradient discussed in **Appendix A**, where the satellites are given kicks at each time step proportional to the area to mass ratio in a direction opposite to the relative velocity between the satellite and the gas. The gas drag is only important while the satellites are small, where type I interactions dominate the reduction in eccentricity, orbital inclinations, and semimajor axis for larger satellites. After Papaloizou and Larwood (2000), type I interactions are simulated with added accelerations $\vec{a}_{\rm I}=-\vec{v}/\tau_{\rm I}$, and $\vec{a}_{\rm damp}=-2\left(\vec{v}\cdot\vec{r}\right)\vec{r}/(r^2\tau_{\rm e})-2\left(\vec{v}\cdot\hat{k}\right)\hat{k}/(r^2\tau_{\rm i})$, where $\hat{k}$ is a unit vector perpendicular to the plane of the disk and $\vec{v}$ is the satellite velocity. The timescale $\tau_{\rm I}$ for type I reduction of the semimajor axis is related to the timescale $\tau_{\rm e}\approx1.4\tau_{\rm i}$ for the damping of the eccentricity and inclination by (Tanaka and Ward, 2004)

$$\tau_I = \frac{1}{C_a \Omega_K}\left(\frac{M_J}{m_s}\right)\left(\frac{M_J}{r^2\sigma_g}\right)\left(\frac{H}{r}\right)^2 = \frac{C_e}{C_a}\frac{\tau_e}{(H/r)^2} \qquad [13]$$

where C_a and C_e are constants of order unity.

The solid inflow is represented by adding orbiting masses with random positions for $r < r_c$ at a rate proportional to F_{in}/f. Constraints on CPU time for the calculations forced the added objects to be 50–600 km in radius – far larger than the small objects expected to follow the gas during the inflow. But once in orbit, the rapid accretion times imply that the particles would coagulate into sizes that are decoupled from the gas in times less than the viscous spreading time of the gas and to accumulate into larger objects on timescales short compared to gas drag timescale τ_{GD}. If these assumptions are valid, starting with 50–600 km objects to accommodate the computational constraints should have little effect on the outcomes for satellites much larger than these sizes. Collisions within the disk are treated as inelastic mergers.

Since the satellites continue to accrete objects after they reach mass m_{crit}, their migration into Jupiter is hastened, and the value of M_T/M_J given by eqn [12] is an upper bound. This is illustrated in **Figure 1** for a constant and uniform flux density F_{in} ($\gamma_{in} = 0$), $\tau_G = 5 \times 10^6$ years, $r_c = 30R_J$, and $\alpha/f = 10^{-6}$, 5×10^{-5}, 5×10^{-4}. The solid lines indicate the oscillating values of M_T/M_J resulting from the occasional loss of a satellite into Jupiter as a function of time normalized by the accretion timescale, and the dotted lines are the values of M_T/M_J from eqn [12], both for the indicated values of α/f. Generally, the total mass of solids delivered to the disk will exceed M_T, so that several satellite systems with $M_T/M_J = 2 \times 10^{-4}$ would be provided with the earlier ones being lost to Jupiter.

Figure 2 shows that for time-dependent influx rates with $\tau_{in} \sim 1.5 \times 10^6$ years, outer radius for flow deposition $r_o = 30R_P$,

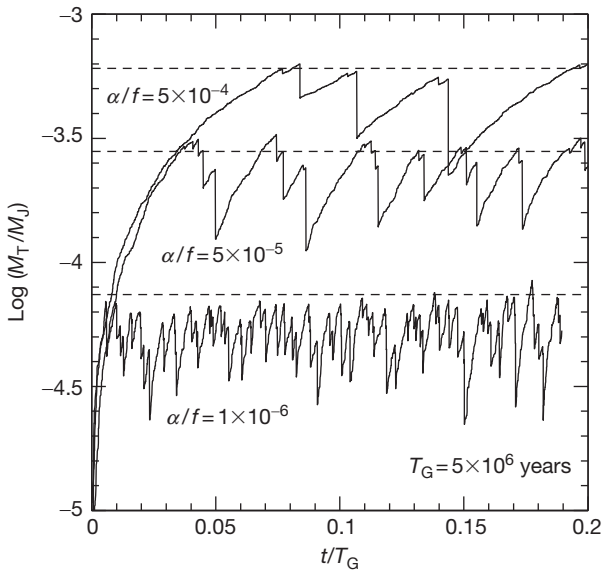


Figure 1 Total mass of satellites as a function of time for a constant rate of inflow for three values of α/f. Modified from Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441, 834–839, with permission.

$\gamma_{in} = 0$, and $\alpha/f = 6.5 \times 10^{-5}$ with $\alpha = 6.5 \times 10^{-3}$, a system of satellites resembling the Galilean satellites is produced with $M_T/M_J \sim 3 \times 10^{-4}$. Recall that M_T/M_J is almost independent of τ_G (eqn [12]) but that disk temperatures are affected by F_{in} (with effective disk temperature $T_{eff} \propto F_{in}^{1/4}$ (Canup and Ward, 2002)). Early in the simulation when the influx is high, the masses of the satellites are given by eqns [11] and [12], but the satellites accreted will be rocky. The properties of the disk must evolve to the point where the midplane temperature of the disk is below 200 K at least beyond the point in the disk where Ganymede accretes a significant fraction of its mass and the accretion of at least Callisto has a timescale $> 10^5$ years to accommodate the latter's possible lack of full differentiation. This timescale requirement generally persists even if Callisto turns out to be differentiated (McKinnon, 1997; Peale, 1999), since influx to the disk and hence the maximum accretion rate are limited by the low disk temperature constraint needed for ice at Ganymede. As the influx wanes, the disk temperature is expected to drop and the component of ice in the satellites is increased. A final generation of Jovian satellites from solar composition inflow with $\tau_{in} = 10^6$ years would form within a disk having temperatures below 200 K exterior to 15 R_J with $\alpha =$ a few $\times 10^{-3}$, a disk opacity $\kappa = O(0.1)\,\mathrm{cm^2\,g^{-1}}$, and a planet temperature $T_J = 500$ K (Canup and Ward, 2002). The constraint of the outer satellites having a large icy component is thereby satisfied. The corresponding timescale for the accretion of the final generation of satellites is $\sim\tau_{in} \sim 10^6$ years, which is consistent with only partial differentiation according to the criteria of Stevenson et al. (1986) and Barr and Canup (2008). Neither the ratio of the total mass of the satellites to the planetary mass, M_T/M_J, nor the distribution of the final satellites is affected much for $-1 < \gamma_{in} < 1.8$, so the results are given in Canup and Ward for $\gamma_{in} = 0$. However, whether most of the mass is deposited close to the planet or far from the planet should affect the distribution of the ice fraction, but this effect of γ_{in} has not been explored. In addition, larger values for γ_{in} may reduce the number of final large satellites (Ogihara and Ida, 2012).

In reality, the satellites are likely to be trapped into orbital resonances sometimes through differential migration of the forming satellites. Such captures were common in the integrations of Canup and Ward (2006) when there were larger outer satellites catching up with smaller inner satellites through type I drift (**Appendix A**). Peale and Lee (2002) demonstrated an assembly of the Galilean satellites into the set of 2:1 MMRs that comprise the present Laplacian relation, where Ganymede is assumed to migrate on the 10^5-year timescale of Canup and Ward and Io and Europa migrate on timescales that are double this value. All the current properties of the resonant system are produced in this scenario, which also requires significant tidal torque from Jupiter to prevent disintegration of the resonant structure from dissipation in Io. However, the assumption of the relative migration rates is ad hoc, and it remains to be shown that detailed plausible disk properties including the opacity as a function of temperature and thereby position in the disk (**Appendix A**), assumed viscosity, etc. can lead to relative migration rates that are consistent with the hypothesis. Still, during the satellite accretion, it is relatively certain that inner small satellites will be trapped into orbital resonances with outer larger satellites that are migrating faster. Such

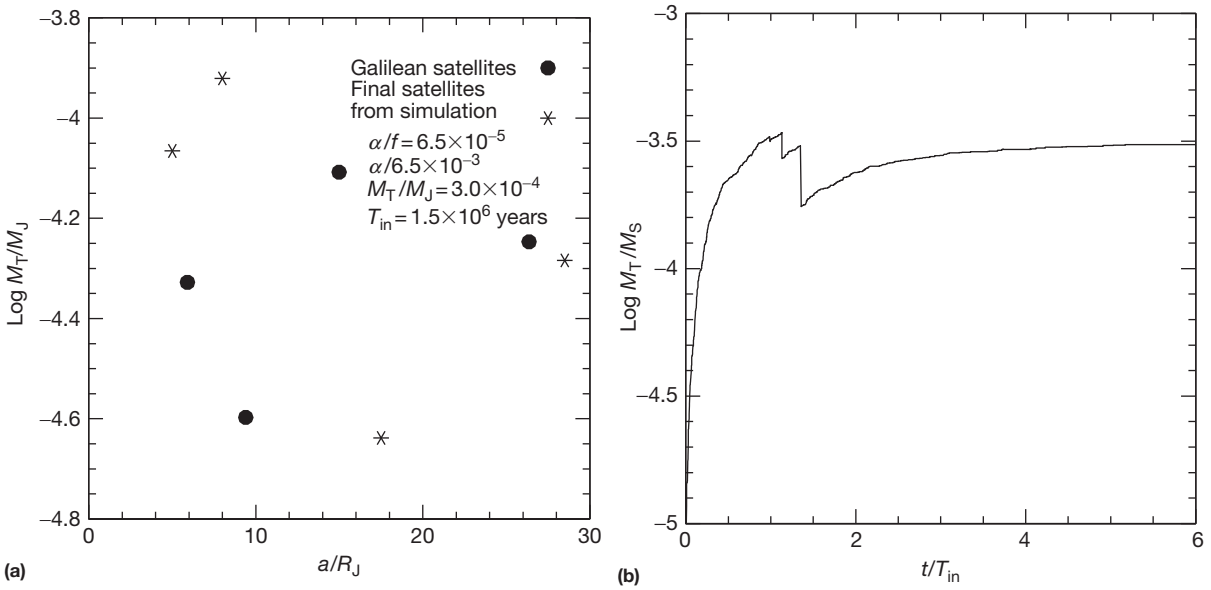


Figure 2 An example of a Jupiter-like system for a waning influx of material onto the disk. (a) Comparison of the masses and locations of the real Galilean satellites with the final set of satellites of a Jupiter-like system from a simulation of a waning influx of material. (b) The time dependence of M_T/M_J for the simulation resulting in the final set of satellites shown in (a). Modified from Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441, 834–839, with permission.

trapping in resonances will hasten the demise of the small satellites interior to larger satellites, where such effects are naturally accounted for in the simulations of the mutual gravitational interactions during differential migration.

The Canup–Ward model successfully produces systems similar to the Galilean satellites (Figure 2). Its virtues include the following: (1) The model is formulated within a plausible model of a waning accretion of the planets, naturally producing not only the observed satellite–planet mass ratios for Jupiter but also those for Saturn and Uranus. (But see Sections 10.17.5 and 10.17.6 for possible caveats.) (2) The waning accretion causes the disk to evolve to sufficiently low disk temperatures to allow water ice condensation and incorporation of that ice into the satellites. (3) The main conclusions are insensitive to the inflow rate and only weakly dependent on the notoriously uncertain value of α. (4) The last generation of Jupiter's satellites accretes sufficiently slowly to allow possible incomplete differentiation of Callisto. (The latter constraint could be relaxed somewhat if it turns out that Callisto is in fact completely differentiated (McKinnon, 1997; Peale, 1999).) (5) The theory is clearly formulated in terms of parameters whose values can be adjusted and the consequences ascertained as new observations and theory provide more constraints. (6) Numerical simulations are consistent with the analytic approximations used in estimating the critical satellite mass attained before loss to the planet and for the total mass fraction M_T/M_J. (7) The simulations produce satellite systems similar to those observed for Jupiter, Saturn, and Uranus for plausible values of the parameters.

Possible caveats to these conclusions include the following: (1) The model requires that the planet contracts to a scale smaller than the satellite system while the nebula is still present, so that an accretion disk forms. While this appears consistent with current contraction models (e.g., Papaloizou and Nelson, 2005; Ward and Canup, 2010), the alternative picture in which a disk is created by spinout from a planet that contracts after the nebula has dissipated would likely produce a very different outcome. (2) A disk with viscosity α values comparable to those inferred for circumsolar disks is assumed. This seems consistent with what is needed to explain the accretion of mass to the planet, given the relatively high specific angular momentum expected for the inflowing gas. But quantitative models justifying such α values in this environment are lacking. Although the effect of the value of α on the conclusions is suppressed by small fractional exponents, the range leading to systems consistent with those of the gaseous planets still implies relatively large values of α. What values of α can be expected from the influx rate as a function of time or from other means of generating turbulence such as stirring by the satellite embryos themselves or MHD-generated turbulence? (3) The model assumes that the outer edge of the inflow region corresponds roughly to the observed scale of the outermost regular satellites. Determination of r_c from hydrodynamic simulations is very challenging, due to the required high resolution, the need to account for the tendency of the circumplanetary disk to viscously spread, and the dependence of results on the assumed nebular scale height. But on balance, current models suggest that the inflow region extended somewhat beyond the current large outer satellites (e.g., Ward and Canup, 2010). Are there processes that would have inhibited satellite accretion in the outer disk? Or could r_c be set by another process, such as the radial extent of a midplane dead zone where MHD turbulence is inactive (Lubow and Martin, 2013)? (4) The assumption that all of the solids are delivered with the gas is dependent on the view that much of the solid mass is in sizes much less than a meter and that small particles do penetrate the gap. An abundance of small particles is predicted by current 'pebble accretion' models, although their

delivery to the disk in the presence of the gap should be more thoroughly investigated. How does a realistic estimate of the solid particle size distribution affect the solid inventory of the disk? Are there viable alternative schemes for solid delivery? (5) The critical dependence on disk properties on the opacity affects in particular the inferred disk thermal profile for a given inflow rate and how type I migration rate varies with r. It is possible that the latter could affect the relative growth rates of satellites at various distances. The Canup–Ward models do not include a time-dependent opacity nor a disk that may evolve from optically thick to optically thin. Has the assumption of a constant opacity seriously altered the evolution?

Sasaki et al. (2010) considered a Canup and Ward-type model, only with two important differences. First, they terminate the accretion shortly after Jupiter opens a gap in the heliocentric nebula, which leaves a hot massive disk, which they assume is quickly reduced by a short viscous spreading timescale. This leaves the satellites that are currently in the disk to interact with the remaining gas. The properties of the disk and satellites therein are now 'frozen' at the time of gap formation, and the disk is now closed. In essence, this marries one of the main premises of the SEMM model (an abrupt end to the inflow upon gap opening) to an actively supplied disk model. A unique and crucial assumption in their model is the creation of an inner cavity in the disk through magnetic coupling to the rapidly rotating Jupiter that extends out to the synchronous orbit (2.25 R_J assumed). This inner gap is motivated from inferences about the differences between classical and weak-line T Tauri stars, where the former's properties are assumed to result from formation of an inner cavity created by the magnetic field of the star. An eccentricity barrier at the inner disk edge (Ogihara et al., 2010) is proposed as the means of stopping the satellites at the boundary. Here, a satellite crossing the boundary because of its eccentric orbit forced by a MMR experiences a positive torque because of a tailwind near apoapse but has no negative torque at the periapse because it is not in the disk on this part of its orbit. There are at least two caveats for this hypothesis: (1) Sakaki et al. adopted the timescales obtained by Ogihara et al. (2010) for type I drift in an isothermal disk. These timescales are appropriate to opposing inside and outside torques. These type I drift timescales are applied to all the satellites in the resonances – even to the satellite at the disk's inner edge. But the outer Lindblad resonances are still in the disk for that satellite and there are no opposing torques from the inside, so the satellite will experience a continued type I torque forcing it toward Jupiter that is now enhanced because there are no interior Lindblad resonances contributing an opposite torque. The timescale for the type I drift is thereby reduced for the satellite at the disk edge from that which was assumed. A possible solution could be obtained by considering the opposing action of the corotation torque at the outer edge of the disk cavity (Massett et al., 2006). (2) Type I timescales are derived for a circular satellite orbit, but the forced eccentricity of the satellite at the disk edge introduces additional resonances in the disk which contribute to the torque. A complete treatment of the disk edge dynamics must be completed before one can have confidence that the eccentricity trap will work.

If the barrier does exist, subsequent satellites are now trapped in a series of MMRs and migration of the whole stack of satellites is resisted by the cavity barrier holding the innermost satellite in place. Once the gas infall terminates, the disk is depleted by viscous spreading and the disk temperature decreases exponentially with time. When the mass of the trapped, resonantly bound satellites exceeds the remaining mass in the disk, the barrier is breached and the innermost satellite is forced into Jupiter presumably still in the MMR with the outer satellites. The next satellite in line then is halted at the barrier but may be pushed into Jupiter as the remaining satellites continue to accrete mass from the disk. The ice content of the outer satellites increases as the disk cools and the ice line moves in. When the gas content of the disk is sufficiently depleted, migration stops leaving a system comparable to the current Galilean satellites.

When Sasaki et al. considered an abrupt cessation of inflow to the disk, their outermost large satellite typically forms too quickly to be consistent with an incompletely differentiated Callisto. However, by assuming a more gradual reduction in the infall, they found a $\sim 10^6$-year accretion time for the outermost body. Callisto accretes slowly in the outer part of the disk cool enough to accrete a major fraction of ice and migrates so slowly that it is often not trapped into a resonance.

Sasaki et al. modeled satellite accretion using a semianalytic procedure developed by Ida and Lin (2004, 2008) to model planet formation. Ogihara and Ida (2012) considered the same physical conditions as Sasaki et al. but modeled satellite accretion with an N-body simulation. They found similar results for Jupiter-like systems, including the finding that a Galilean-like configuration in which several satellites are captured into mutual 2:1 resonances is a very common outcome when an inner disk cavity is included with the eccentricity trap.

There are several other aspects of the Sasaki et al. (2010) proposal that are controversial: (1) The assumption that accretion is halted at the time of gap formation is not consistent with the global 3-D calculations of Bate et al. (2003), who found accretion is not halted until $M \sim 10 M_J$. (2) There is no analysis of the disk ionization needed to establish the magnetic coupling to the gas that creates the inner cavity. (3) It is not clear that the large magnetic field necessary to effect the disk standoff is actually to be expected. On the other hand, the existence of a cavity at some point may well be required to explain the current subcritical rotation rates of Jupiter and Saturn (Takata and Stevenson, 1996). (4) And their model assumes a satellite accretion timescale that increases with orbit radius, whereas in an inflow-regulated disk, this timescale decreases with increasing r. This appears likely to have affected the architecture of their final satellite systems, particularly in the case of Saturn-like systems without an inner cavity (see the discussion in the next section).

The Canup–Ward (2002, 2006) model is a reasonably complete and plausible model of satellite formation within a model of the latter stages of the accretion of Jupiter, albeit not without caveats. It is consistent with the composition and positional constraints and produces the gross features of the satellite systems of the giant planets for plausible choices of the controlling parameters, which differ from planet to planet. Some choices of the parameters yield systems comparable with the observed Galilean satellites. The vertical accretion onto the disk is consistent with the high-resolution 3-D simulations. The accretion of the satellites in the circumplanetary

disk is specifically treated as a function of disk properties that change as the accretion winds down, where plausible timescales for the various processes are assigned that are consistent with the observed properties of the satellites. The encouraging conclusions from the model depend on the assumed value of α, the presence of sufficient small solids that can be efficiently delivered to the disk by the gas, and the disk's thermal model (and therefore the assumed opacity). Refinement of the model should develop realistic constraints on these parameters.

10.17.5 Saturn System

The Saturn system is far different from the Jupiter system, and aside from the conclusion that the regular satellites formed in a cold, dissipative disk, the details of the formation of the satellites therein are less understood. The Saturn system is dominated by a single satellite, Titan, which contains 96% of the combined mass of all Saturn's satellites. Whereas Jupiter's satellite densities decrease progressively with increasing distance from the planet, at Saturn, they are strikingly nonmonotonic: the densities of the satellites Mimas, Enceladus, Tethys, Dione, Rhea, Titan, Hyperion, and Iapetus are 1.15, 1.61, 0.97, 1.48, 1.23, 1.88, 0.57, and 1.08 g cm^{-3}, respectively (e.g., Jaumann et al., 2009). These densities imply a substantial ice content for all the satellites, but in common with Jupiter, the fraction of the planet's mass contained in the satellites is near 2×10^{-4}. Interior to these satellites are Saturn's main rings, which together contain about the mass of a Mimas-sized satellite, and multiple smaller ringmoons. The ringmoons are dynamically young (e.g., Esposito, 1986; Goldreich and Tremaine, 1982) and may be recent products of the collisional expansion of ring material beyond the Roche limit (Charnoz et al., 2010). The small mass of the satellites other than Titan, the lack of any rock-dominated satellites inside Titan's orbit, the nonmonotonic distribution of densities, and the origin of the rings comprise the major mysteries of the formation of Saturn's satellite system.

While the low density of Hyperion (mean radius $\langle R \rangle \sim 150$ km) is significantly affected by porosity, this is expected to have a minor effect on the densities of larger Tethys ($R = 530$ km) and Iapetus ($R = 718$ km). In particular Tethys' density is the lowest of all outer solar system objects having $R > 500$ km and, for plausible densities for rock and nonporous ice, implies a rock content between 0% and 6% (e.g., Thomas, 2010). The main rings are also more than 90–95% pure water ice (e.g., Cuzzi et al., 2010). These are extremely ice-rich compositions compared to a bulk solar composition of roughly half rock/half ice. In contrast, the mass fractions of rock in Enceladus, Dione, and Titan are similar to those of Ganymede and Callisto. The current high density of Enceladus may reflect loss of water from the observed plumes at the south pole (Porco et al., 2006). The loss rate of neutral H_2O is estimated to be ~75–1600 kg s^{-1} (e.g., Fleshman et al., 2010). The history of activity on Enceladus is totally unknown, but if current activity has persisted for a significant fraction of Enceladus' lifetime, the satellite could have lost a significant fraction of its ice and started with a more Mimas-like density.

Titan is about 50% water as deduced from its mean density. Estimates of its moment of inertia from Cassini data suggest that its interior may be incompletely differentiated like Callisto (Iess et al., 2010), although this result assumes hydrostatic equilibrium and is therefore uncertain due to the possible contribution of nonhydrostatic structures (e.g., Gao and Stevenson, 2012). The mole fraction of methane in Titan's atmosphere is about 5% in the lower atmosphere and the rest is essentially all N_2 except for traces of argon, a tentative detection of neon, and small amounts of heavier hydrocarbons (Niemann et al., 2005, 2010). Aside from a mole fraction of $2.1 \pm 0.8 \times 10^{-7}$ for ^{36}Ar and $2.8 \pm 2.1 \times 10^{-7}$ for ^{22}Ne, no other primordial inert gases have been detected (Niemann et al., 2010).

If the N_2 in Titan's atmosphere were trapped in the icy planetesimals that went into Titan as direct condensation from the gaseous phase, or if it were trapped in amorphous ice or as a clathrate hydrate, the noble gases would have also been trapped. The low upper bound on the noble gases implies then that the N_2 came into Titan in the form of NH_3 (Niemann et al., 2005, and references therein). The nondetection of krypton and xenon led to several suggestions for the origin of the atmosphere consistent with the lack of these gases. One possibility is that the icy planetesimals formed at temperatures >75 K so that the noble gases and CH_4 were not captured, but NH_3 and CO_2 were. However, Owen and Niemann (2009) showed that the detection of Ar but not Kr or Xe could instead be due to the limited sensitivity of the Huygens probe instrument given the relative abundance of these elements in other solar system bodies, making such arguments inconclusive.

Heavier hydrocarbons are thought to be products of photochemistry initiated by photodissociation of methane (e.g., Niemann et al., 2005). The relatively short timescale for the survival of CH_4 from photodissociation of 10–100 million years (Wilson and Atreya, 2004; Yung et al., 1984), together with the lack of any isotope fractionation in the carbon of CH_4 like that seen in nitrogen and oxygen, means that CH_4 must be continuously supplied from Titan's interior (Niemann et al., 2005). This scenario may require the continuous supply of CH_4 to Titan's atmosphere to be created through chemistry within Titan (Zolotov et al., 2005). We are certainly left with the constraint that the temperature eventually becomes cool enough in the Saturn disk to allow water ice to persist but perhaps not so cold as to permit stable methane clathrates.

Most models of the Saturnian protosatellite disk invoke a surface density, midplane temperature, and pressure that increase with decreasing r as in the Jupiter examples in the preceding text. This would normally favor the formation of rocky satellites close to the planet with icy satellites further out, yet the inner satellites are ice-rich at Saturn. For an accreting disk, the rate of heat generated by viscous dissipation is proportional to the inflow rate (Canup and Ward, 2002), and so for very slow rates of inflow, ice can be stable in the inner disk. In the limit that viscous dissipation is unimportant and the planet's luminosity is the dominant heat source in the disk, the ice stability distance is within three Saturnian radii once the planet's temperature falls below about 500 K (e.g., Canup, 2010). A cold, solar composition inner disk would produce ice-rock satellites, rather than extremely ice-rich objects like Mimas and Tethys. Furthermore, a smooth disk surface density profile consistent with Titan's mass would generally yield much more massive satellites in the inner disk than those observed. Thus, the inner Saturnian satellites are not easily explained by a standard disk model.

The Mosqueira–Estrada SEMM model for the Saturnian disk (2003a) starts with a smoothed minimum mass nebula based on the total satellite system mass, rather than on the specific distribution of mass in the current inner satellites. The temperature and surface mass densities are assumed to vary as $1/r$ as in the Jupiter model. As in the Jovian SEMM disk, a steep surface density gradient is assumed just outside the centrifugal radius defined earlier, which for Saturn is just outside the orbit of Titan. We argued earlier that it is improbable that such a steep gradient could have been maintained during the disk's formation, when the viscosity required to transport angular momentum outward so that the planet can accrete mass (e.g., Estrada et al., 2009) would have smoothed out such variations (e.g., Ward and Canup, 2010). But interestingly, capture of Hyperion into its current 4:3 MMR with Titan seems to require such a steep density gradient ($\sigma_g \sim r^{-3}$) if Hyperion formed as Titan formed and was brought into the resonance by gas drag, as seems likely (Lee and Peale, 2000). As Hyperion nears its final mass, which was probably considerably larger than its current mass because pieces chipped off Hyperion typically do not remain in the resonance and are not reaccreted, gas drag overwhelmingly selects the 3:2 MMR preventing migration to the 4:3 if the surface mass density decreases less rapidly than r^{-3} (Lee and Peale, 2000).

In the SEMM model, Titan forms in the inner, dense disk in 10^4–10^5 years and would likely be fully differentiated (Mosqueira and Estrada, 2003a). Melting of ice occurs in Titan's interior if it forms in $\lesssim 10^6$ years even in the limit that there is no radiogenic heating and the satellite forms from small impactors that deposit their energy close to the surface (Barr et al., 2010), so that substantially more rapid formation implies differentiation. In the SEMM model, Titan is saved from loss due to inward type I migration by opening a gap in the very low viscosity disk, with the disk then removed due to nonviscous processes (e.g., photoevaporation) so that Titan is not lost by type II migration. Mosqueira and Estrada form Iapetus in their outer, low surface mass density disk, where it migrates inward by type I drift to the point where this torque vanishes because of the steep density gradient between the inner and outer disks. The ice-rich composition of Iapetus is then ascribed to the preferential delivery of water to the outer disk by the ablation of differentiated heliocentric planetesimals, because objects passing through the low-density outer disk become hot enough to lose their ice but not their rock (Mosqueira et al., 2010).

To account for the ice-rich compositions of the inner moons, Mosqueira and Estrada (2003b) suggested that the inner Saturnian disk was initially too warm for ice and that early rocky embryos that formed there were lost as they spiraled into Saturn. Subsequently, as the disk cooled, the ice condensed to form ice-rich satellites about 10^5 years after the end of Saturn's accretion when the disk gas surface density had decreased to $\sim 10^4$ g cm^{-2}. However, the survival of the inner moons would then require a very rapid removal of the inner gas disk just after they formed; otherwise, they too would be lost to gas drag in the order of a few $\times 10^4$ years (for Mimas) to $\sim 10^5$ years (for Rhea). Mosqueira and Estrada (2003b) qualitatively discussed seven different reasons for why there might be no large satellites interior to Titan. Their preferred explanation is that the tidal torque from a fully formed Titan removed the inner disk before other large inner satellites could form, but how a tidal torque causes the inner material to flow rapidly inward is not explained. For a disk with no inflow, satellite growth would generally proceed more rapidly in the inner disk due to shorter orbital periods and a higher disk surface density, implying that if the inner disk contained a mass much larger than that in the current inner moons (and this is the case for the SEMM disk), satellites much larger than the current inner Saturnian moons would likely form before Titan formed.

An underlying assumption of the SEMM or any minimum mass model is that the satellites form after the planet has finished accreting gas, so that there is no inflow supply to the disk. However, the period of inflow to the disk is long (comparable to the lifetime of the nebula), while the timescale for solids orbiting in the disk to accrete into satellites is short. Thus, it seems more probable that Saturn's satellites formed throughout the lifetime of its disk as it was being supplied by an inflow, rather than waiting until after the inflow had ended, similar to the conditions advocated for the Jupiter system. The Canup–Ward model (2002, 2006, 2009), as well the simulations of Sasaki et al. (2010) and Ogihara and Ida (2012), adopt this premise.

Canup and Ward (2006) simulated the accretion of Saturn's satellites from an actively supplied disk with an N-body model as described in **Section 10.17.4**. Jovian-like systems with four or five similarly sized large satellites were the most common outcome in the Canup and Ward simulations. But depending on the stochastic timing of when the inflow ends relative to the oscillations of satellite growth and loss seen in **Figure 1**, other system structures were also produced. **Figure 3** shows a simulation with $r_c = 30R_S$, $\gamma_{in} = 0$, $\alpha/f = 6 \times 10^{-5}$, and $\alpha = 6 \times 10^{-3}$ that produced a Saturn-like system with one large satellite and numerous small satellites (Canup and Ward, 2006). Initially, four large satellites formed, but the inner three are lost to collision with the planet as the inflow wanes. After the last large inner satellite is lost, the final inflow delivers enough mass to produce Dione- and Rhea-sized satellites in the vacated inner region of the disk. During this period, the inflow rate is so slow that the ice stability line moves inward to a few Saturnian radii. When the infall ends and the gas disk dissipates, a seven-satellite system remains with $M_T/M_S = 1.8 \times 10^{-4}$ and the most massive satellite at $\sim 14.6R_S$ that contains 70% of the mass of the total satellite system mass. Its close orbital spacing to the satellite at $11.3R_S$ suggests a future collision may occur. There are two satellites bracketing the position of Iapetus but with larger orbital eccentricities. The inner satellites in this simulation are sufficiently close together that collisions are likely with a reduction in the total number of satellites.

The Canup and Ward simulations were the first to show that a smooth disk could produce approximate analogs to Saturn's satellite system. However, systems with only one large satellite were not common, and they did not produce any systems in which one satellite contained $>70\%$ of the final satellite system mass. Although their calculations produced the small, ice-rich satellites inside Titan's orbit, they could not account for the large increase in the ice fraction of these satellites over the half rock/half ice content of the influx nor could they account for the nonmonotonic distribution of the rock–ice content of the inner satellites. This difficulty has motivated an alternative scenario described in the succeeding text that

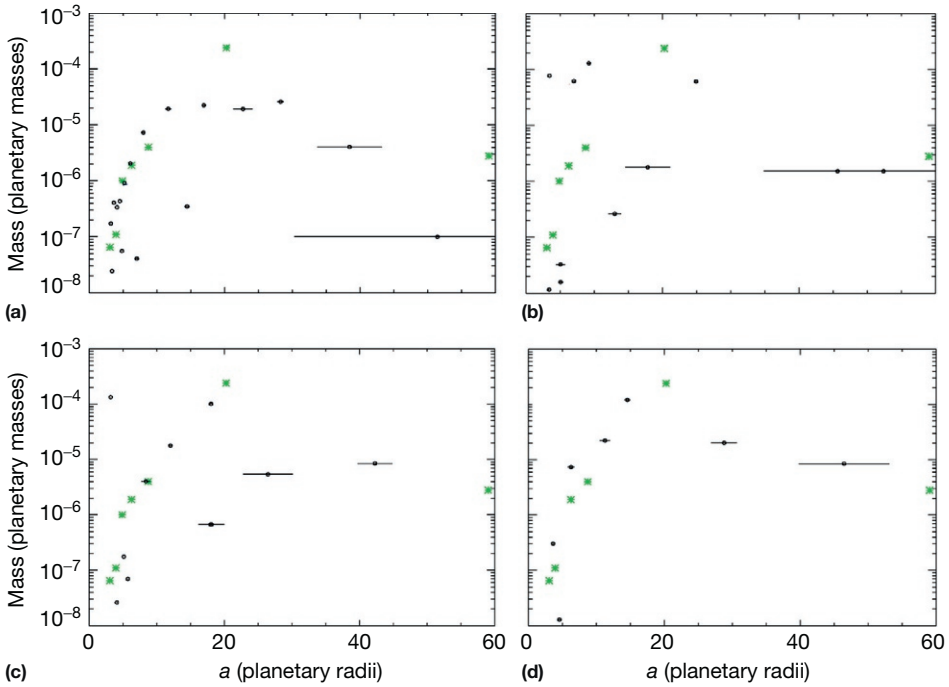


Figure 3 Results of a satellite accretion simulation from Canup and Ward (2006) that produced a Saturn-like system of satellites. Here, the inflow to the disk is assumed to decrease exponentially with time with a time constant τ_{in}. Black circles show the simulated satellites (with horizontal lines proportional to orbital eccentricities); Saturn's satellites are shown as green stars. (a) Multiple satellites form as solid material flows into the disk ($t = 0.2\tau_{in}$). Once satellites grow to $M \sim$ few $\times 10^{-5}$ planet masses, they begin to migrate inward. (b) At $t = 1.4\tau_{in}$, the system resembles Jupiter's Galilean system, with four similarly sized large satellites. The inner three large satellites are lost to collision with Saturn, with the last large satellite lost at time $t = 3\tau_{in}$ (c), when the inflow rate has slowed substantially. The final system (d) at $t = 10\tau_{in}$ has a single large Titan-like satellite orbiting at 15 planetary radii.

produces nearly pure ice inner satellites as well as a nearly pure ice ring system.

Sasaki et al. (2010) simulated the accretion of satellites using a semianalytic, Monte Carlo model. This model closely reproduced the oscillations in the total satellite system mass seen in **Figure 1**. However, when considering the same conditions as Canup and Ward (an inflow that decayed exponentially with time and a planet without an inner cavity in its disk), they found the most common result was a Saturn-like system with a single large satellite. Ogihara and Ida (2012) performed N-body simulations of the conditions considered in Sasaki et al. and found Jupiter-like systems were the most common result for disks without inner cavities, similar to the Canup and Ward (2006) findings. All three works use the same expressions for the rate of type I migration and the quasi steady-state gas surface density. However, their treatment of the satellite growth rate differs. Sasaki et al. used a semianalytic procedure in which they use a satellite accretion timescale that is proportional to the local orbital period, leading to $\tau_{acc} \propto r^{5/4}$ (their eqn [13]) and a longer growth timescale for outer satellites compared to inner ones. But in an inflow-supplied disk, the rate of satellite growth is controlled by the rate of delivery of material to the disk rather than by the local orbital period (Canup and Ward, 2002). For a uniform inflow per area, this leads to $\tau_{acc} \propto 1/r^{4/5}$ (Canup and Ward, 2006) or $\tau_{acc} \propto 1/r^{2}$ (Ogihara and Ida, 2012), depending on how the width of a satellite's feeding zone is defined. In both the Canup and Ward (2006) and Ogihara and Ida (2012) simulations, satellites in the outer disk thus have shorter growth timescales than those in the inner disk.

By imposing a slower accretion in the outer disk, the Sasaki et al. model would tend to artificially delay the growth of outer Titan-sized objects in the time it takes an inner Titan-sized object to migrate inward. This appears to have biased their results in favor of single satellite systems. Ogihara and Ida (2012) suggested that a radially nonuniform inflow, in particular one that is centrally concentrated, might be able to produce a similar result, although this has yet to be demonstrated.

Can the very ice-rich compositions of the inner moons be explained in the context of such models? Canup (2010) proposed that as the last of the large satellites was lost, tidal stripping from its outer layers created Saturn's rings and ultimately the inner satellites out to Tethys. This scenario requires that this last satellite be differentiated as it crossed the Roche limit with the outer layer being nearly pure ice or an ice-covered ocean. This possibility is supported by Canup (2010) who notes that accretion should bring a Titan like body to near the melting point of ice even on the relatively long accretion time scale of 10^6 years expected under the waning influx of material to the disk. A sustained eccentricity $\sim 2 \times 10^{-3}$ for $k_2/Q \sim O(10^{-2})$, which is a balance between excitation by perturbations of other satellites and damping by tidal dissipation in the satellite, would provide somewhat more than the latent heat of melting an ice shell in the $\sim 10^4$ years the satellite spends near the Roche limit before colliding with Saturn. This heat input plus that from accretion is not

unlikely to produce the needed differentiation into an ice water mantle over a rocky core.

When the satellite's orbit decays within the Roche limit set by its mean density (located at $\sim 1.75R_S$ for a Titan-density satellite, where R_S is Saturn's current mean radius), tidal distortion begins to remove material from its ice shell. The removal of ice causes the satellite's mean density to increase until the remnant satellite is marginally stable. This continues until either the remnant collides with the planet or its higher-density rocky core disrupts. The Roche limit for rock is at $a_{R,rock} \sim 1.5R_S$. Planet contraction models predict that the newly formed Saturn's radius would have been between $R_P = 1.5R_S$ and $R_P = 1.7R_S$ (e.g., Fortney et al., 2007; Marley et al., 2007). For $a_{R,rock} \leq R_P$, the remnant satellite collides with the planet before its core disrupts, and tidal stripping produces a pure ice ring.

Canup (2010) used SPH to simulate a differentiated Titan-sized satellite as it spirals inward from its initial Roche limit. The total mass of ice produced via tidal stripping depends on the location of the planet's surface: In the limit that $a_{R,rock} = R_P$, $\sim 10^{25}$ g of ice is stripped into orbits exterior to the satellite ($\sim$10% of the original satellite's mass), while if $a_{R,rock} < R_P$, less ice is produced before the remnant satellite hits the planet. As each ringlet of ice is stripped, strong shepherding torques from the remnant satellite rapidly repel it, driving exterior material into orbits beyond the Roche limit for ice in 10^2 years. Here, the stripped material accumulates in a pure ice moon whose final mass is $m_o \leq 10^{25}$ g. Canup estimated that the location of synchronous orbit around a newly formed Saturn would be located at about $3R_S$ due to the slower rotation of the more distended planet, so that this ice moon eventually spirals inward due to planetary tides, but at a slower rate than the original satellite due to its smaller mass. For planetary tidal parameters $(k_2/Q)_P$ (with likely values $10^{-6} \leq k_2/Q_P \leq 10^{-5}$; e.g., Charnoz et al., 2009), the ice moon takes $\sim 10^6$ years $[5 \times 10^{-6}/(k_2/Q)_P]$ $[10^{25}\, g/m_o]$ to decay within the ice Roche limit, where it disrupts into a massive ice ring. This timescale is long enough that the ring forms after the remnant satellite has collided with Saturn and when the circumplanetary gas disk has essentially dissipated. Mosqueira and Estrada (2011) claimed that such a ring would be destroyed by gas drag, because its surface density decreases with time (and with it, the ring's lifetime against gas drag) as the ring spreads. But a ring formed at the very end of gas accretion can survive, because there is also a decrease in the gas surface density with time both before the ring forms and as the ring evolves due to the waning inflow, which they did not consider. However, their point is a good one for any rings formed by satellites that decayed within the Roche limit earlier in a planet's gas accretion.

The Canup model implies an initial ring that is a factor of 20–500 times more massive than Saturn's current rings. In contrast, prior ring origin models produce an initial ring comparable in mass to the current rings (Charnoz et al., 2009; Dones, 1991). Salmon et al. (2010) found that such dense rings can evolve to something resembling Saturn's ring by viscous spreading after $O(10^8)$ years, but it is not clear how the ignored MMRs with the closest satellites such as Mimas would affect this evolution. As a ring viscously spreads, it spawns satellites from its outer edge (e.g., Charnoz et al., 2010), and Canup estimated that for such a massive ring, the first spawned satellite would have a mass comparable to that of Tethys, whereas Charnoz et al. (2011) found the first mass spawned to be comparable to Rhea's mass. Spawned moons evolve outward due to first resonant interactions with the ring and then by tidal evolution, with the mass of spawned moons decreasing with time as the mass of the ring decreases due to viscous spreading (Charnoz et al., 2010, 2011). Mimas, Enceladus, and Tethys would be by-products of the formation of Saturn's rings. However, the dynamics of such satellite orbit expansions have not been investigated (see succeeding text).

Extending this idea, Charnoz et al. (2011) proposed that all of the satellites interior to and including Rhea were spawned from an initial ring containing a few $\times 10^{24}$ to few $\times 10^{25}$ g. Dione and Rhea contain $\sim 10^{24}$ g in rock, much greater than the total rock in the inner moons ($\sim 10^{23}$ g) that could be delivered by external bombardment during the LHB (Canup, 2013). The Charnoz et al. (2011) model thus requires an initial ring with a substantial rock component. To leave the rings as essentially pure ice, they creatively suggested that the rock in the rings was initially in the form of large chunks that were able to open gaps in the ring and undergo type II migration. A direct simulation of a Rhea-sized chunk embedded in a ring is used to show that the chunk is expelled from the ring's outer edge, where it would then form the core of a spawned satellite. It is, however, not clear why the ring's rock would be contained in fragments many orders of magnitude larger than the accompanying ice particles, since the strength of rock and ice is similar and because ice and rock particles would likely form rock–ice aggregates of varying mean densities, nor whether a simulation that included a large number of rock chunks would be able to remove them all from the ring with nearly perfect efficiency. A second condition of the model is that Dione and Rhea evolve out to their current orbits through tides, which requires an extremely small tidal Q parameter for Saturn of $Q < 2000$. It is generally believed that the Q is of order 10^4–10^5 for gas giants. Lainey et al. (2012) considered astrometric observations of Saturn's satellites over the last century and found $Q \sim 1700$. This result is controversial, both because it implies a much different Q for Saturn than similar techniques predicted for Jupiter (Lainey et al., 2009) and because their data also showed Mimas migrating inward rapidly, whereas it should be migrating outward due to both tides and ring torques.

There are several important caveats to both the Canup (2010) and Charnoz et al. (2011) models. First, neither can produce the nonmonotonic distribution of satellite compositions without invoking ad hoc degrees of freedom. In Charnoz et al., a pure ice initial Tethys is proposed to result from stochastic variation in how rock is delivered from the initial ring. The Canup model implies an essentially pure ice initial Tethys but then requires additional processes to explain Enceladus' rock content while avoiding the subsequent accretion of too much rock onto Tethys. Next, can the initial satellites accreted outside the Roche radius retreat far enough before the next one is spawned to allow the minimum spacing of the satellites, or is an additional sequence of mergers required? The simulations of Charnoz et al. (2011) show that typically more inner moons are formed than are currently in the system, although they utilize a simplified accretion model. Crida and

Charnoz (2012) used an analytic merger model and predicted a mass–distance distribution of satellites in good agreement with that seen at Saturn. However, their accretion model does not include gravitational interactions between the moons, including MMRs.

The most critical constraint comes from the limited tidal evolution of the current satellites. Slow differential expansion of the orbits will lead to a plethora of MMRs that will be encountered, and capture within will lead to eccentricity growth that will destabilize the system. That the tidal evolution of the system has been limited is implied by calculations of evolution of the Saturn system that brings Enceladus to the 2:1 MMR with Dione (Peale, unpublished). Reducing the satellite semimajor axes to start the integrations consistent with their mutual tidal evolution brings Tethys and Dione and, almost at the same epoch, Mimas and Enceladus inside 3:2 MMRs. The integrations had to start outside these resonances to avoid the certain captures therein, which implies that the evolution of the satellite system is limited over the 4.6×10^9-year history to start outside these resonances. There is no obvious means to escape from these resonances and the eccentricity growth therein leads to crossing orbits before the resonances become unstable. The implied lower bound on the average Q of Saturn exceeds 90 000, which is higher than that inferred from the proximity of Mimas to Saturn. The system configuration is such that Enceladus' semimajor axis must start outside 0.9832 its current value if the two 3:2 MMRs are to be avoided.

One way MMRs among the satellites spawned by the massive icy ring in the Canup and Charnoz et al. models can be traversed without capture is to note that the initial orbital expansion of these satellites as they interact with a massive ring will be very rapid compared to tidal evolution. If the orbit of the inner satellite expands the width of the MMR in a time less than the libration period of the resonance variable, capture can be avoided. It is possible that the inner moons (Mimas, Enceladus, and Tethys) achieved nearly their current positions through the ring interactions, and subsequent tidal evolution can be severely limited. But this possibility has not yet been demonstrated. In the Charnoz et al. (2011) model, Dione and Rhea become too distant to experience ring torques and they must migrate over vast distances at the much slower tidal rates. If 90 000 is the real lower limit on the historical tidal Q, such distant tidal migration is unlikely. If Q is much lower than this and the extensive migration could ensue, the capture of Dione and Rhea and the inner satellites into destabilizing MMRs is likely. Details of the orbital evolution of the satellites under the variety of possible constraints discussed have not been investigated.

The challenge of explaining the variation in densities among the mid-sized Saturnian satellites has motivated alternative origin scenarios that form these satellites as intact ejected clumps from large impacts. If the impacting objects are differentiated, such clumps can have compositions that vary significantly from those of the parent objects. Sekine and Genda (2012) considered the collision between a smaller satellite and a Titan-sized body as the latter migrated into the inner disk due to type I migration. Asphaug and Reufer (2013) proposed that an initial multiple moon system at Saturn was destabilized through mutual collisions that ultimately left only Titan and that the current mid-sized moons were clumps ejected from such collisions. Both works use SPH simulations to demonstrate that such collisions can produce large intact fragments having similar sizes and a wide range of rock fractions as observed among Saturn's mid-sized moons. However, both works face a common challenge: how to avoid having the clumps ejected from an impact reaccrete onto the target object. Ejecta are initially on crossing orbits with the target and would generally be rapidly reaccreted. If ejecta are in the form of small particles, then their mutual collision times can be shorter than the reaccretion timescale, and in this case, they can collisionally damp to a ring, which could potentially accrete into separate stable satellites. But this process would also seem likely to remix the compositions of the clumps and seems unlikely to yield multiple satellites with well-separated orbits as observed.

We conclude our discussion of Saturn with Iapetus, which presents a special problem in that its orbit is nearly circular, but inclined by about 7.49° relative to the local Laplacian plane. The latter inclination would imply a capture origin, but because eccentricity damping due to the tidal dissipation in Iapetus is very weak, a capture origin is inconsistent with its current small orbital eccentricity ($e_I = 0.0283$). The Laplacian plane is that plane on which a satellite's orbit precesses at nearly a constant rate with nearly a constant orbital inclination. The Laplacian plane nearly coincides with the equatorial plane for close satellites due to torque from the planetary oblateness, but it coincides nearly with the planet's orbital plane for distant satellites due to the solar torque. An extensive debris disk settling to this plane would therefore be warped as the local Laplacian plane changed from equatorial plane to orbital plane with distance from the planet (e.g., Ward, 1981). The regular satellites accumulated from debris disks at distances where the oblateness of the planet dominated the location of the Laplacian plane. But at Iapetus' distance from Saturn, the local Laplacian plane is inclined by 15° to Saturn's equator, where the ascending nodes on the orbital plane of Saturn's equator (obliquity = 26.7°) and the Laplacian plane coincide. Can Iapetus have accumulated in the debris disk as implied by its small orbital eccentricity yet be inclined to the local Laplacian plane by 7.49° today?

Ward (1981) offered a tidy affirmative answer by pointing out that the debris disk itself contributes to the precession of objects within it. The Laplacian plane thereby would have a lower inclination to Saturn's equatorial plane at the time the disk was present, and Iapetus could have accreted there as assumed in the previously mentioned discussion. Iapetus' orbital plane precesses with a current period of about 3000 years (Ward, 1981), and this period would have been shorter when the disk was present. If the disk could be dispersed on a timescale short compared to the orbit precession period at the time, the Laplacian plane would rotate toward its current position and leave Iapetus with its observed low eccentricity and 7.49° orbital inclination. The short dispersal time necessary, $O(100)$ years, is not consistent with accretion times of the satellites of order 10^3–10^4 years in the theories discussed in the preceding text or for allowing sufficient time for rocky satellites to spiral into Saturn in the inner disk, leading to the formation of icy satellites there. Viscous dissipation of the disk on this short timescale requires values of $\alpha = O(1)$ (Ward, 1981), which is implausible. Because of the difficulties imposed by such rapid disk dispersal, the anomalous orbit of

Iapetus is still a puzzle. A capture origin seems unlikely, because sufficient gas drag to circularize the orbit would also have damped the inclination to the local Laplacian plane.

Although the structure of the circumplanetary disk is still poorly constrained and there are uncertainties about the nature of the solid content in the gas inflow, the Canup–Ward model is perhaps a first step toward an understanding of the origin of Saturn's satellite system, since a system with no large satellite interior to Titan appears attainable. The loss of large inner satellites to type I migration provides a means to produce a pure ice initial ring and ice-rich inner satellites. The loss of silicate embryos while the nebula was too warm to allow water ice condensation followed by cooling to allow the remnant ice to accrete onto the small satellites is an attractive alternative (Mosqueira and Estrada, 2003a,b). But there have been no detailed disk evolution models to show that the sequence of events needed for this scenario can occur. Explaining the distribution of densities in the inner satellites remains a challenge, and current ideas ultimately depend on stochastic events, which are difficult to model with great confidence due to inherent uncertainties in the relevant processes. We are much closer to understanding the details of the origin of Jupiter's satellites than we are to understanding the origin of the Saturn system.

10.17.6 Uranus System

The Uranus system of satellites is Jupiter-like in the sense that it contains a few large satellites with $M_T/M_U \sim 1 \times 10^{-4}$ (M_T is the total mass of the satellites). All the satellites out to Oberon with the exception of Miranda have nearly equatorial orbits with densities–radii of the major satellites Miranda, Ariel, Umbriel, Titania, and Oberon of 1.2/236, 1.7/579, 1.4/585, 1.7/789, and 1.6/761 ($g\,cm^{-3}\,km^{-1}$), respectively. Inside Miranda's orbit are 13 small satellites that are the most densely packed in the solar system and are thereby unstable and experience relatively frequent collisions (French and Showalter, 2012). Thirteen remnant rings in the inner system are testimony to the frequent destruction and reassembly of the satellites in the inner system. The Canup–Ward (2006) model can produce a Uranus-like system of the major satellites, where the lower rate of inflow and cooler disk temperatures about the smaller mass planet produce satellites with a mixture of rock and ice that is consistent with the densities. However, these authors note the difficulty of moving the disk material from heliocentric orbit to the equatorial plane of Uranus when the obliquity of Uranus is 98°. On the other hand, if the Uranus obliquity were small when the satellites formed and only later tipped over by an adiabatic process, the Canup–Ward model would be immediately applicable. But with more certainty than the Saturn case, Uranus will not open a gap in the solar nebula, and the influx may have a different distribution over the disk than that appropriate for Jupiter. So, the Canup–Ward model may have to be modified.

Accounting for the maintenance of a coplanar, equatorial system of satellites about Uranus as the planet gains its large obliquity is the major obstacle to understanding their origin. An adiabatic scheme for tilting Uranus, where equatorial orbits would be maintained by adiabatic invariance (e.g., Goldreich, 1965), requires a resonance between the precession rates of the spin axis about the orbit normal and the precession of the orbit about its local Laplacian plane. This requires the spin precession rate to be increased drastically from its current value. Boué and Laskar (2010) accomplished this by assuming that Uranus originally had a massive satellite with an orbital radius of about 0.01 AU. However, Morbidelli et al. (2012) pointed out that the Laplacian plane for such a satellite is close to Uranus' orbital plane and this satellite orbit would not follow the equator as the Uranus obliquity increased. Because of its large mass, this satellite would also hinder at least Oberon and Titania from doing so. Morbidelli et al. concluded that a sudden tilting of Uranus with a giant impact is the only remaining viable means of accounting for the large obliquity, a scheme which had been proposed before (e.g., Slattery et al., 1992).

The most viable timing of the impact is after the planet has completed its gas accretion and thereby already possessed a satellite system. The impact would likely disrupt this system into crossing orbits where collisions would disintegrate the satellites into a debris disk. The latter scenario was verified numerically by Morbidelli et al. (2012), so these authors consider the evolution of a protosatellite disk after the massive collision that tilts Uranus modeled as gravitationally interacting concentric rings. After the tilt, the disk is still in the former equatorial plane, but Uranus' new equatorial plane corresponds to the new 98° obliquity. The rotational oblateness of Uranus, parameterized by the gravitational harmonic coefficient J_2, causes differential precession of the ascending nodes of the disk rings (or of particles in the disk) so the disk becomes a thick torus spanning the equatorial plane. Collisions within the torus damp the ring to the new equatorial plane, where the new disk can accrete into equatorial satellites. However, the J_2 effect competes with the self-gravity of the disk, where the latter dominates for those portions of the disk that are further from Uranus and cause the disk to precess as a unit while maintaining its original inclination to the equatorial plane. Such dominance of self-gravity of the particles was first realized for the epsilon ring of Uranus, where the periapse longitudes precessed together maintaining the ring coherence (Goldreich and Tremaine, 1979).

If the current value of J_2 is assumed, the nodes are randomized by differential precession only slightly beyond Miranda's orbit near 0.001 AU. Beyond Miranda's distance, the inclined ring would lead to satellites highly inclined to the equatorial plane. But existing simulations indicate that the impact should create an equatorial disk near 0.01–$0.03M_U$ extending out to $3R_U$, where R_U is Uranus' radius (Slattery et al., 1992). This equatorial ring enhances J_2 sufficiently to extend the randomization of the node longitudes of the rings comprising the disk out beyond Oberon's orbital distance. The mutual collisions damp the disk down to the equatorial plane where the equatorial satellites can be formed. But if the tilt to the current obliquity is from 0° to 98° with the single impact, the satellite orbits will be equatorial but retrograde! On the other hand, if Uranus had a preimpact obliquity >30°, the probability of prograde orbits exceeds 40% (Morbidelli et al., 2012). This complicates the scenario somewhat as it requires Uranus to have had at least two giant impacts, but such impacts near the end of planetary accretion are thought to have been common (Morbidelli et al., 2012). The equatorial disk formed from the impact probably could

not contribute significantly to the major satellites as they are too far away, but contribution to the small inner satellites is possible.

In contrast to the opinion expressed in Section 10.17.4 against starting with a pristine disk for the formation of the regular satellites of Jupiter, here is an example of the creation of a disk around a giant planet from which the satellites containing about 10^{-4} of the Uranus mass could be formed. If the solar nebula had already been removed at the time of impact, there would be no further accretion onto the disk after the disk was formed. If the debris disk from the collisional breakup of the original satellites contained little gas, loss of solids from gas drag would be eliminated, and the total satellite mass would then be $O(10^{-4}\,M_U)$ appropriate for the Canup–Ward model of the original accretion.

An alternative to this scheme has been proposed by Crida and Charnoz (2012). Here, the satellites are created from a massive disk mostly inside the Roche radius that was created at the time of the giant impact that tilted Uranus. The satellites are created one by one just outside the Roche radius as the disk spreads viscously, and they migrate away from the disk from disk torques initially and eventually through tidal expansion of the orbits. This idea has the appealing features that only a single impact is necessary to create a system that will be equatorial and prograde. However, the details of its execution pose some severe problems. First, there most likely would have been an accretion disk about Uranus as it was being formed, and satellites would have formed in this disk. Since Uranus must have been mostly completely formed at the time of the impact, it must have had a preexisting satellite system in its equatorial plane as assumed by Morbidelli et al. (2012). This preexisting system is assumed not to exist in the Crida and Charnoz scheme. Next, the disk formed by the impact is inside about $3R_U$ so that Oberon must be moved from this distance to almost $23R_U$ if it is spawned from the ring. Even if the tidal dissipation function Q for Uranus were sufficiently small to accomplish this feat over the age of the solar system, Titania, more massive than Oberon and at a smaller orbital distance, would have moved out more rapidly. Finally, the disk produced by a Uranus-tilting impact would initially be vapor, which would evolve differently than a disk of solid particles as considered by Crida and Charnoz. It is difficult to see how the Crida–Charnoz scheme could be the origin of the Uranian satellites.

The Slattery (1992) simulation of the giant impact contained only 5000 particles in the proto-Uranus and 3000 particles in a $2M_E$ impactor. Rather substantial changes in the outcome resulted from SPH calculations of the lunar-forming impact on the Earth as the resolution was increased substantially above that of earlier calculations (Canup, 2004b), so it is important to do simulations of the Uranus impact with the highest resolutions currently practical to see if conditions assumed (e.g., the mass and extent of the equatorial disk generated by the impact) in the successful scheme for forming the prograde, equatorial satellites by Morbidelli et al. actually prevail.

10.17.7 Neptune System

A plausible theory has emerged for the current configuration and content of the Neptune system of satellites. Here, the task is simpler than for the other gas planet systems, because Triton's destruction of the original system during its capture has erased all concern for accretion disks, gaps, solar nebula interactions, etc. The dominant characteristic of the Neptune satellite system is the existence of the large satellite Triton (NI) ($R = 1353$ km), in a close ($r \approx 14.3R_N$, $R_N = 24\,766$ km) circular, retrograde orbit (obliquity 156.8°). The remaining regular satellites are prograde. Neptune also has relatively few known satellites compared with the other major planets, and all but two of those, Triton and Nereid (NII), were unknown until the Voyager spacecraft observations (Smith et al., 1989) and subsequent ground-based discoveries (Holman et al., 2004). Nereid's orbital eccentricity of 0.75 brings it no closer than about 1.4×10^6 km from Neptune's center because of its extremely large semimajor axis of 5.51×10^6 km ($222.6R_N$) – well outside Triton's distance.

The massive satellite Triton is blamed for most of the features of this unusual system. The retrograde orbit means that Triton was almost certainly captured intact from heliocentric orbit. It has been proposed that Triton was captured in a three-body exchange where Pluto was expelled from orbit around Neptune (Farinella et al., 1980), that it was captured by gas drag (Mckinnon, 1984; McKinnon and Leith, 1995), or that it collided with a satellite already in orbit about Neptune with a mass a few percent of Triton's mass (Goldreich et al., 1989). The first hypothesis suffers from the likely multiple subsequent close encounters of the ejected Pluto with Neptune and essentially no means of leaving Pluto in the currently stable 3:2 MMR with Neptune, whereas the second has low probability. A collision would have been sufficient to capture Triton into a very eccentric orbit extending a significant fraction toward the Hill sphere boundary at $r_H \approx (M_N/3M_\odot)^{1/3} a_N \approx 4.5 \times 10^3 R_N$ (a_N = heliocentric distance) while not destroying it.

An alternative method of capture of Triton involving the disruption of a binary KBO with characteristics similar to the Pluto–Charon binary pair has been proposed (Agnor and Hamilton, 2004, 2006). Binary systems among the small bodies in the solar system appear to be ubiquitous with ~16% of near-Earth asteroids (Margot et al., 2002), ~2% of main belt asteroids (Merline et al., 2002), and ~11% of KBOs (Stephens and Noll, 2006) having companions. The fraction of small bodies in binary pairs is probably larger than these such that binary–planet encounters are likely. When a binary pair comes sufficiently close to a planet such that the Hill sphere of the pair in the planetary gravitational field is smaller than their separation, they become tidally unbound. Depending on the phase of the relative motion in the binary and the inclination of the orbital plane with that of the encounter trajectory, one of the planetesimals can be left in bound orbit about the planet while the other escapes. A simulation with binary pair masses $m_1 = m_T$ and $m_2 = 0.1m_T$ separated by $20R_1$ in circular orbit, where m_T is the mass of Triton and R_1 is the radius of m_1, closest approach distance $8R_N$ well inside the tidal disruption radius $r_{td} = 26R_N$ (R_N = Neptune radius), resulted in 95% of the binaries being disrupted for $0 < v_\infty < 2.0$ km s^{-1}, ~50% of the encounters resulting in capture of either m_1 or m_2 for $v_\infty < 0.35$ km s^{-1}, and ~50% of the encounters resulting in capture of m_2 for $0.35 \lesssim v_\infty \lesssim 1.55$ km s^{-1}. For the latter velocity interval, capture of m_1 is still possible but rare.

Although either collisional capture (Goldreich et al., 1989) or binary exchange capture (Agnor and Hamilton, 2004, 2006)

can yield reasonable orbits that can damp to the observed retrograde circular orbit observed today, Agnor and Hamilton show that with at least 11% of KBOs being binary systems, capture by collision has negligibly small probability compared with the binary exchange capture. If the original satellite system of Neptune were similar to that of Uranus, collision of Triton with any satellites this large would have destroyed it, making the currently inclined orbit of a recollected Triton impossible (Ćuk and Gladman, 2005). So, it is almost certain that Triton was captured gently without such a risk, although a collisional capture remains an improbable possibility. Nogueira et al. (2011) verified the arguments for the capture of Triton by the disruption of a binary KBO, and Vokrouhlicky et al. (2008) argued that the capture must have occurred early in solar system history in the context of the Nice model of this history (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005).

The properties of a prograde satellite system that likely existed around Neptune prior to Triton's capture are of course unknown. An original satellite system like that of Uranus extending from about $5R_N$ to beyond $20R_N$ would have been perturbed into crossing orbits, and the satellite system would be thereby destroyed. During the times when Triton's periapse was inside $20R_N$, energy would be removed from Triton's orbit impulsively at each passage through the disk of material leading to circularization times of 10^4–10^5 years while Triton cleared most of the material from the disk outside of about $5R_N$ (Ćuk and Gladman, 2005).

If interactions with such a disk were not sufficient to circularize Triton's orbit, subsequent tidal dissipation would suffice. With $k_T(=k_2)=0.1$ and $Q_T=100$, Goldreich et al. (1989) found that an initial Triton orbit with semimajor axis $a_T \approx 10^3 R_N$ would damp to nearly its current circular orbit with $a_T \sim 14.3R_N$ from tidal dissipation in Triton in about $4–5\times10^8$ years – comfortably shorter than the age of the solar system. However, the large amount of dissipation probably melted a significant fraction of the satellite, which drastically hastens tidal evolution (Mckinnon et al., 1995; Shock and McKinnon, 1993). The Love number k_2 would then be $O(1)$ (closer to that of a fluid sphere) instead of $O(0.1)$ and $Q_T \sim O(10)$ instead of $O(100)$, leading to a factor of $\sim$100 shorter eccentricity decay timescales.

We can estimate the time required to damp Triton's eccentricity (e_T) with the following procedure. During the damping of Triton's eccentricity, it probably existed in an asymptotic rotation state defined by the zero of the averaged tidal torque given, for example, by eqn [B.17] in **Appendix B**. We determine de/dt from eqn [B.27], where we assume $\dot{\psi}_s$ is the asymptotic value given as a function of e by the zero of the third component of eqn [B.17]. If we also assume that the periapse position is maintained at Triton's current distance, then the semimajor axis can be represented as $a_0/(1-e)$ where a_0 is the current value. With these assumptions, the time to decrease Triton's eccentricity from 0.9 to 0.01 is $3.3\times10^5\ Q_T/k_T$ years. The tides raised on Neptune by Triton will have a negligible effect (Banfield and Murray, 1992; Goldreich and Soter, 1966). With the Goldreich et al. choices of k_T and Q_T, the time to damp Triton's eccentricity is 3.3×10^8 years, which is the several hundred million years damping time found by Goldreich et al. With the more likely value of k_T/Q_T, this time is only 3.3×10^6 years.

But tidal dissipation may have been less efficient in the circularization of Triton's original eccentric orbit than inferred by Goldreich et al. because solar perturbations cause periodic increases in periapse distances (Benner and McKinnon, 1995). The Goldreich et al. analysis ignored these Kozai oscillations in eccentricity (see **Section 10.17.9**), which would have occurred if the original maximum orbital inclination was comparable to the current one. These 1000-year variations in e_T cause the mean periapse distance to be much larger than that obtained by Goldreich et al. with the consequence that the tidal damping time, with the same values of Q_T and k_T, and expression for dissipation used by Goldreich et al. would yield a timescale of 3.5 Gy for the eccentricity reduction (Ćuk and Gladman, 2005). Reducing this number to 35 My with the more likely values of Q_T and k_T in the preceding text still makes the tides adequate.

Consistent with the capture scenario and subsequent melting, Triton was found to have a young and active surface. The observed, currently active plumes are almost certainly solar-driven (Smith et al., 1989), but the geologically young volcanic plains and cantaloupe terrain are clearly endogenic, where such surface modification may be ongoing (Shenk and Jackson, 1993; Stern and McKinnon, 2000). Triton's density is about 2 g cm^{-3}, which indicates some ice content. Triton's original constituents were obtained directly from the solar nebula at its location of formation, but there may be significant contamination with debris from the original satellite system. The volatile content was probably reduced during the rather extreme tidal heating during the eccentricity damping, although some N_2 was retained from the early dense atmosphere and most of the water is still there (Lunine and Nolan, 1992). Triton and Pluto are similar in size and density, but Pluto may have lost a lot of volatiles from the impact that created Charon (see **Section 10.17.8**), and one cannot infer that this loss is the same as that from Triton. The current density does not constrain the place in the solar nebula where Triton may have been formed. But an initial composition similar to that of Halley's comet is consistent with observations of the surface ices, where the severe reduction of CO relative to N_2 and CO_2 can be understood by hydrothermal processing resulting from the tidal heating (Shock and McKinnon, 1993).

While Triton's orbit was extended, any regular satellites inside about $5R_N$ would also have been disrupted and destroyed from the periodic dips of the periapse inside $5–8R_N$, depending on the initial semimajor axis. There are likely to have been several satellites inside the periapse distance of Triton in comparison with the regular systems of the other major planets. Orbit eccentricities would have been limited to maximum values near 0.3, since the rate of decrease of the eccentricity by tidal dissipation within the satellite exceeds Triton's ability to increase it for larger values (Banfield and Murray, 1992). Two satellites with semimajor axes of $3R_N$ and $5R_N$ would have overlapping orbits that would persist in overlapping for times at least of the order of the eccentricity damping time of 10^5–10^7 years. A collision between any of Neptune's current inner satellites with its neighbor would lead to their mutual destruction. This follows from the relation (Stevenson et al., 1986)

$$\frac{1}{2}m_i v_i^2 \sim m_s S + \frac{3}{5}\frac{Gm_s^2}{\gamma R_s} \qquad [14]$$

where m_i and m_s are the masses of the impactor and satellite, respectively; v_i is the relative velocity at impact; R_s is the satellite radius; $S \sim 10^6$ erg g^{-1} is the material binding energy; and $\gamma \sim 0.1$ is a factor introduced to account for the inefficiency in converting the impact kinetic energy into kinetic energy of the fragments. The impactor kinetic energy must exceed the energy stored in material strength plus the self-gravitational energy by a sufficient amount to break up the body. From eqn [14] and eccentricities of 0.3, Naiad (NIII) could destroy all of the satellites except Proteus (NVIII), and any of the satellites NIV to NVII could destroy Proteus. This implies that the current satellite system could not have existed prior to Triton's orbit circularization (Banfield and Murray, 1992). The debris from the first generation of satellites would settle into the equatorial plane in circular orbits and recollect into a second generation of inner satellites with nearly circular orbits and zero inclinations, where all memory of Triton's perturbations would be thereby lost. Had any of these inner satellites survived destruction, their orbits would have been left inclined to Neptune's equatorial plane. The lack of such inclined orbits among the inner satellites must mean that there are no survivors among those early inner satellites.

The inner satellites observed today are not those formed after the circularization of Triton's orbit. The flux of comets through the system would have destroyed all of the current inner satellites except for Proteus (1989N1) over the lifetime of the solar system (Colwell and Esposito, 1992; Smith et al., 1989). The debris from such a destructive collision would settle to the equatorial plane where the satellites can reaccrete. The rings of small particles are likely to be remnants of these collisional breakups. Any distant irregular satellites of Neptune (Holman et al., 2004) would also have their orbits perturbed drastically by Triton if they were captured before or during Triton's capture and orbital decay and their numbers depleted (Ćuk and Gladman, 2005).

10.17.8 Pluto System

Pluto and its large satellite Charon orbit about their center of mass with a period of 6.3872273(3) days at a separation of 19,573(2) km with zero eccentricity (Buie et al., 2012a). (The numbers in parentheses denote the 1σ uncertainty in the last digits.) The rotation of both bodies is synchronous with the orbital motion, which makes this system the only one in the solar system known to have reached the end point of tidal evolution. Like the Earth–Moon system, the Pluto system has a high specific angular momentum that favors the formation of Charon by a large oblique impact (Mckinnon, 1984, 1989; Stern et al., 1997). Canup (2005) had performed SPH simulations of impacts on the Pluto precursor with impactor mass a significant fraction of the target mass with a variety of compositions and differentiation states. The most probable outcome of a simulation that is successful in placing a sufficient amount of mass in orbit outside the Roche distance is for Charon to be launched as an almost intact satellite in a highly eccentric orbit. In these examples, the impactor is undifferentiated and Charon is composed almost entirely of material from the impactor, which results from its being sheared off in the oblique impact. Some of the discussion of the giant impact creating the Moon in Section 10.17.2 is applicable to Pluto–Charon, but the impact associated with the latter involves much less energy. There is little heating of Charon, little water ice is vaporized, but there is probably sufficient energy deposited in Pluto at impact coupled with radiogenic heat to cause its differentiation (Canup, 2005). An impact thus appears a quite natural mechanism to produce Charon, assuming that large collisions were probable in the early Kuiper Belt. However, the impact does not, as it turns out, provide a straightforward explanation for Pluto's tiny satellites discovered in recent years. The objects present a current puzzle, which we describe in some detail later in the text. The gravitational collapse scenario pointed out earlier in the text for widely spaced, massive Kuiper Belt binaries probably is not applicable to the Pluto system because of the difficulty of assembling a sufficient amount of mass into a gravitationally unstable swarm (Nesvorny et al., 2010).

Pluto has four known additional satellites in nearly circular orbits that are nearly coplanar with the orbit of Charon. The first two discovered (Weaver et al., 2006) are named Nix and Hydra. Best-fit Kepler orbits by Buie et al. (2012b) yield periods, semimajor axes, and eccentricities of 24.85494(6) d, 38.20163(7) d; 48 831(25) km, 64 736(11) km; and 0.00232(19), 0.00658(15), respectively. Pluto's fourth satellite, S/2011 (134 340) 1, also known as P4, was discovered on 28 June 2011 (Showalter et al., 2011). The satellite was found in archival images, where the long time span allowed the determination of the following Kepler orbit parameters (Buie et al., 2012b): period = 32.1866(22) d, semimajor axis = 57 890(120) km, and eccentricity = 0.0045(11). The fifth satellite, S/2012 (134 340) 1, also known as P5, was discovered on 14 separate sets of images from the Wide Field Camera on the Hubble Space Telescope on June 26, 27, and 29 and July 7 and 9 (Showalter et al., 2012). P4 is about 1/10 as bright as Nix, and P5 is about 1/2 as bright as P4. The orbital period of P5 is 20.2 ± 0.1 days with orbital semimajor axis near 42 000 km. All of the orbital planes are within $\lesssim 0.15^\circ$ of Charon's orbital plane. With the orbital period of Charon at 6.3872273 days, the ratio of the orbital periods of each of the satellites to that of Charon is 3.16 (P5), 3.89 (Nix), 5.03 (P4), and 5.98 (Hydra) reasonably close to 3:1, 4:1, 5:1, and 6:1, respectively. The parameters of the Pluto system are summarized in Table 1.

Although the orbital periods of Charon, Nix, and Hydra are nearly in the ratios 1:4:6, no resonance variables are likely to be librating at the present time (Lee and Peale, 2006). The small satellites orbit about the center of mass of Pluto and Charon, and that position from the orbit-fitting procedure gives a Charon–Pluto mass ratio of 0.1168(55). Pluto's radius from the mutual event light curves is 1164(22.9) km (Young and Binzel, 1994) and Charon's radius from a stellar occultation is 603.6(1.4) km (Sicardy et al., 2006). With the masses determined by Buie et al., these radii yield densities of 2.03(0.06) and 1.65(0.08) g cm^{-3} for Pluto and Charon, respectively. The physical parameters of Pluto and Charon are consistent with the results of the giant impact origin of the system (Canup, 2005). Lee and Peale (2006) showed that the orbital motions of Nix and Hydra are markedly non-Keplerian because of the large Charon–Pluto mass ratio. As the masses of the new satellites are increased within the range allowed by the uncertainty in the albedos, the close proximity of the 3:2 MMR between Nix and Hydra causes increasing fluctuations in the

Table 1 Pluto system

	Mass (g)	*Radius (km)*	*Rotation period (days)*	*Orbital period (days)*	a *(km)*	*e*
Pluto	1.302×10^{25}	1151	6.38723	6.38723	–	–
Charon	$0.1165 m_P$	603.6	6.38723	6.38723	$19573 = 17.0 R_P$	0
P5	?	5–14[a]	?	20.2	$42000 = 36.5 R_P$	~0
Nix	$O(3 \times 10^{-5}) m_P$[b]	23–68[a]	?	24.855	$48831 = 42.4 R_P$	~0
P4	?	7–20[a]	?	32.2	$57890 = 50.3 R_P$	~0
Hydra	$O(6 \times 10^{-5}) m_P$[b]	30–84[a]	?	38.202	$64736 = 56.2 R_P$	~0

[a]Assumed albedo range (0.04–0.35).

[b]The stability of P4 places upper limits on the masses of Nix and Hydra that are about an order of magnitude less than these estimates from astrometry (Youdin et al. 2012).

eccentricities. Although the maximum masses allowed within these constraints allow possible libration of the 3:2 resonance variable involving the periapse longitude of Hydra, these maximum masses are probably already ruled out because they would also excite Charon's eccentricity beyond the current observational uncertainties (see Buie et al., 2012a). A dynamic fit to a sufficiently dense set of observational data will yield constraints on the masses of Nix and Hydra (Lee and Peale, 2006). The fact that Nix, Hydra, P4, and P5 are in orbits coplanar with Charon's orbit is consistent with their being accreted in a dissipative disk of much smaller particles.

One way to create such a disk is from the debris left over from the giant impact that trapped Charon into its orbit about Pluto. However, Nix and Hydra reside at ~42 and 56 Pluto radii (R_P) from the Pluto–Charon center of mass, and debris from the impact will not extend out to these distances (Canup, 2004a, 2005, 2011). A disk extending beyond the location of the 6:1 mean motion commensurability with the initially close orbit of Charon is likely formed from partially differentiated impactor and target although usually not from impacts of homogeneous bodies (Canup, 2011). If the small satellites accreted from this disk of material, an appealing hypothesis for transporting Nix and Hydra, and presumably P4 and P5 discovered later, was investigated by Ward and Canup (2006).

Charon's orbit must expand from its initial size in the process of reaching the current state of completed tidal evolution of dual synchronous rotation (Cheng et al., 2014). One can infer from the closeness of Nix and Hydra to the 4:1 and 6:1 mean motion commensurabilities with Charon that the satellites may have occupied a MMR at each commensurability. Being so trapped would mean they could be pushed out ahead of Charon as Charon's orbit tidally expanded. But the resonance must be chosen with care for this to happen.

There are a number of resonances that can be occupied at the 4:1 and 6:1 mean motion commensurabilities characterized by the angular arguments of terms in the expansion of the perturbing potential

$$\phi_{ml} = (m+1)\lambda - \lambda_C - (m-l)\varpi - l\varpi_C \qquad [15]$$

where λ and λ_C are the mean longitudes of the satellite and Charon and ϖ and ϖ_C are the longitudes of periapse. The angle ϕ_{ml} is written for mth-order resonances of the form $m+1$:1 with $m=3$ for Nix and $m=5$ for Hydra. The terms in the expansion of the perturbing potential are represented by $\Phi_{ml}(a, a_C, e, e_C) \cos \phi_{ml}$, where terms involving orbital inclinations are ignored and where $\Phi_{ml} \sim e^{m-l} e_C^l$. The only resonances at the $m+1$: 1 mean motion commensurability that can preserve the satellites during Charon's orbital expansion are for $m=l$ involving only the eccentricity of Charon's orbit in the coefficient Φ_{ml}. This follows from the fact that all those resonances involving e and hence ϖ lead to steady growth in e as the orbits expand with eventual instability. The resonance involving only the eccentricity and longitude of periapse of Charon is called a corotation resonance, since the satellite orbits at the same angular velocity as the resonant potential in the expansion of the perturbing potential due to Charon, the so-called pattern speed. In resonances of this type, there are forced eccentricities for Nix and Hydra whose magnitudes vanish with e_C, but these do not lead to crossing orbits or other unstable configurations. The stability of a corotation resonance vanishes as $e_C \to 0$.

Inspection of eqn [B.22] in **Appendix B** and the local text show that tides raised on Pluto can generally either increase or decrease the orbital eccentricity e depending on the values of both rotation rate $\dot{\psi}_P$ and e but that tides on Charon always decrease the eccentricity, since Charon's angular velocity $\dot{\psi}_C$ will always be near the mean orbital motion n. For relatively larger values of $\dot{\psi}_P$, there is a competition between tides on Pluto trying to increase e and those on Charon trying to decrease e. Cheng (2011) and Cheng et al. (2014) had shown that with a judicious choice of the ratio of the rate of dissipation in Charon to that in Pluto, it is possible to keep Charon's e nearly constant near its initial value throughout the tidal evolution from its close proximity to Pluto to its current separation in dual synchronous rotation. Thereby, the condition for maintaining Nix and Hydra in the 4:1 and 6:1 corotation resonances with Charon during the expansion of the latter's orbit can be satisfied, but that is not the end of the story.

There are two difficulties with the Ward and Canup hypothesis. First, in an investigation to test the migration hypothesis for Nix and Hydra, where the two small satellites are initially placed outside the 4:1 and 6:1 commensurabilities and Charon's orbit is tidally expanded, there were no captures into either corotation resonance by itself in thousands of trials (Cheng, 2011; Cheng, Lee and Peale (2014a)). Next, by carefully placing one of the satellites into a corotation resonance, it was possible to transport it to its current position. But the parameter space that permitted the migration of one satellite in a corotation resonance does not overlap with that which permitted the other to be so transported (Lithwick and Wu, 2008). It is not possible to transport both Nix and Hydra within corotation resonances. A possible alternative was found by Cheng (2011) where in many cases, a test particle was trapped into more than one resonance at the 5:1, 6:1, 7:1, and 8:1 commensurabilities during thousands of trials with the

gravitational coefficient $J_2 = 0$ for Pluto. Interestingly, when caught in the multiple resonances at a given commensurability, the evolution proceeded to the final state without e growing to instability. All the resonance variables stopped librating at the end of the evolution, but the test particle was often left with a significant eccentricity. There was one example where all of the resonance variables librated about 0° with Charon's and the test particle's orbit aligned that proceeded to the current configuration for Hydra near the 6:1 commensurability with both eccentricities damping to zero. There was no example of stable transport in multiresonances at the 4:1 commensurability.

Unfortunately, this mechanism for transporting Nix and Hydra in multiple MMRs has a fatal flaw. A condition for occupying several resonances at a mean motion commensurability is that the orbits be aligned or antialigned with $\langle \dot{\varpi}_C \rangle = \langle \dot{\varpi} \rangle$. This was possible in the examples because $J_2 = 0$ was assumed. But the rapidly rotating Pluto, softened by the impact, will have a large hydrostatic J_2 (initially >0.1) leading to $\dot{\varpi}_C \gg \dot{\varpi}$. The conditions for multiresonance transport of the small satellites will not prevail.

Lithwick and Wu (private communication) successfully transported Nix and Hydra in three-body resonances (with $J_2 = 0$) that were so delicate that their assembly is very unlikely by natural processes and the presence of P4 and P5 and the finite value of J_2 are major obstacles. Cheng, Peale and Lee (2014b) integrated more than a million sets of initial conditions distributed over eccentricities, semimajor axes, and true anomalies with Nix at a range of initial positions relative to Charon but always starting outside the 4:1 mean motion commensurability. Hydra was always just outside the 3:2 mean motion commensurability with Nix or within one of the resonances at that commensurability. Finite but minimal masses of Nix and Hydra were assumed with Charon's albedo. There were no survivors for either $J_2 = 0$ or $J_2 =$ the hydrostatic value. The frustrating failure of the resonance transport hypothesis to place the small satellites, formed simultaneously with Charon at the time of the impact, at their current positions relative to Charon has inspired alternative suggestions.

Establishing a disk of material after the Pluto–Charon evolution is complete has been suggested by Lithwick and Wu (2008). Such a disk, if sufficiently dissipative, will settle to the plane of the Pluto–Charon orbit. To produce satellites orbiting in the same direction as Charon, the disk would need to have a negative specific angular momentum in the frame rotating with Pluto's orbital motion, since Pluto–Charon have a retrograde orbit. Existing studies of the stability of the Pluto–Charon phase space establish an unstable region spanning Charon's orbit. Stern et al. (1994) found $0.47a_C \lesssim a \lesssim 2.0a_C$ and $0.65a_C \lesssim a \lesssim 2.15a_C$ to be unstable for prograde and retrograde orbits, respectively. Pires dos Santos et al. (2011) found that Nix and Hydra decrease the stability of the region beyond $2a_C$ and that the only stable region between Nix and Hydra is near the 5:1 mean motion commensurability with Charon, where we find P4. The unstable region includes the 2:1 and 3:1 mean motion commensurabilities, and we find P5 just outside the 3:1. It would appear that the small satellites may have selected only the nearest stable regions in the Pluto–Charon phase space. N-body calculations to test this hypothesis have not yet been performed.

A conceivable way of establishing the disk after the evolution of the Pluto–Charon system is a collision of two large objects within Pluto's sphere of influence. At best, this is an event of low probability, but Pires dos Santos et al. (2012) maximized that probability by assuming a heliocentric disk density 1000 times that of the current Kuiper Belt, which would prevail before the orbits of the outer planets were expanded. The probability of collision within the Pluto–Charon sphere of influence is further enhanced by one of the objects in the collision being already in a temporarily captured orbit about Pluto–Charon. With the maximum lifetime of such captured orbits being only $O(100)$ years, Pires de Santos et al. found the time between collisions of objects with sufficient mass to account for Nix and Hydra to be so much longer than this that they abandoned this scenario as a possible origin of a dissipative disk that could have produced the small satellites.

Although it is generally agreed that Charon is the result of a glancing impact on the Pluto precursor with subsequent tidal evolution to the current dual synchronous state, the origin of the small satellites remains uncertain. It seems clear that these satellites formed from a dissipative disk of material, but collisional creation of a disk with sufficient mass that extends throughout the region of the small satellites remains improbable.

10.17.9 Irregular Satellites

The irregular satellites that orbit all of the major planets occupy large orbits with significant inclinations and eccentricities. Recent observations have increased the number of known irregular satellites to 96 (Gladman et al., 1998, 2000, 2001; Holman et al., 2004, http://ssd.jpl.nasa.gov/sat_elem.html, http://ksiezyce.republika.pl/table_en.html). They often extend to distances from their primaries that are close to the stability limits of about half of the planet's Hill sphere radius (Hamilton and Burns, 1991; Hamilton and Krivov, 1997). Similar orbital elements for members of clusters of the irregular satellites imply breakup events after the primary body was captured (e.g., Gladman et al., 2001). The only viable means of establishing these systems of large orbits with high eccentricities and inclinations is through capture from the large number of planetesimals orbiting the Sun early in solar system history (e.g., Burns, 1986). Capture can result from a loss of energy by the planetesimal while it is within the planet's Hill sphere. Such energy loss can be either from gas drag in an extended planetary atmosphere (Pollack et al., 1979), from a collision between two unbound planetesimals leaving one with sufficient energy loss to be bound, or from a collision with a smaller object already in stable orbit about the primary (Colombo and Franklin, 1971). The gas drag mechanism for at least the progenitor of the Himalia group of prograde irregular satellites orbiting Jupiter has been explored in detail by Ćuk and Burns (2004), where it is shown that such capture must occur near the end of a waning gas disk lifetime to avoid losing the satellites to the planet. Gas drag no doubt resulted in the capture of many planetesimals into stable orbits about the major planets, but only the very last ones have survived. The occupancy of orbital resonances by two of the retrograde satellites of Jupiter implies that at least a thin disk was necessarily in place at the time of the capture to allow a weak gas drag to bring the satellites to the resonances (Saha and Tremaine, 1993). An alternative mechanism of capture involves a rapid

growth of the primary, while a planetesimal is temporarily trapped in the Hill sphere, to expand the region of stable orbits (Heppenheimer and Porco, 1977; Saha and Tremaine, 1993). The binary exchange capture for the irregular satellites that is highly probable for the capture of Triton (Agnor and Hamilton, 2004, 2006) is probably not as effective for the capture of the distant irregular satellites, since an irregular satellite is less likely to have ever been close enough to the planet for the binary Hill sphere to have been smaller than the separation of the binary members (Vokrouhlicky et al., 2008).

Still another means of capture of the irregular satellites is a two-body encounter within the Hill sphere, where one of the two participants has sufficient energy loss to remain captured. An interesting variation of nondissipative capture has been proposed by Nesvorny et al. (2007) within the Nice model of early solar system evolution (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005). In the Nice model of solar system evolution, all the major planets were formed between 5 and 20 AU with a sea of planetesimals with total mass between 20 and 35 Earth masses between 20 and 35 AU from the Sun. The initial separation of Jupiter and Saturn is inside the 2:1 MMR. Jupiter migrates inward and Saturn outward from the scattering of the remaining planetesimals in their vicinity such that near 700 million years after formation, Jupiter and Saturn cross the 2:1 MMR in the wrong direction for capture, but Saturn's eccentricity is excited that destabilizes Neptune and Uranus into crossing orbits between themselves and with Saturn. There are repeated planetary encounters, Neptune and/or Uranus is scattered into the sea of planetesimals leading to a vastly increased flux of planetesimals into the inner solar system, and the asteroid belt is disturbed. The flux of objects into the inner solar system results in the late heavy bombardment (LHB) therein. After many close encounters, the orbits of Neptune and Uranus decouple and the eccentricities and inclinations are damped by dynamic friction as the orbits migrate outward by scattering the abundant planetesimals. The current architecture of the outer solar system is thereby produced.

But the many close encounters among Neptune, Uranus, and Saturn within the population of planetesimals led to the capture of planetesimals in stable orbits around these planets. Planetary encounters with Jupiter are rare, so Jupiter's irregular satellites must be captured by some other mechanism, like the gas drag and collisions discussed earlier in the text. In a typical computer run that produces the current architecture, Neptune, Uranus, and Saturn capture many more irregular satellites than are observed with distributions in orbital elements that match the observed distributions remarkably well. So, if the Nice model or some modification of the same that results in many planetary encounters in the presence of planetesimals is close to reality, the Nesvorny et al. scheme must be important in the capture of irregular satellites. Of course, this mechanism of capture only occurs more than 700 million years after the initial formation of the planets, which may mean that irregular satellites captured before this time by, say, the collision or gas drag mechanisms are replaced by the new population – except Jupiter's irregular satellites may be those captured at earlier times. It is interesting that the systems of irregular satellites are so similar if scaled to the Hill sphere of each planet (Jewitt and Sheppard, 2005), when a different mechanism may be important for Jupiter than for the other three giant planets.

The capture of satellites should lead to a random distribution of orbital inclinations relative to the planet's orbital plane (~ecliptic), but there are few irregulars known with inclinations between 47° and 141°. Carruba et al. (2002) gave a lucid explanation of the scarcity of satellites with orbital inclinations within this range. Solar perturbations dominate the variations in the orbits of these satellites, and in an analytic secular theory of the restricted three-body problem involving perturbations of the satellite orbit by the lowest-order solar tidal term, the motion is averaged over the mean motions of both the planet and the satellites. In this approximation, the component of the satellite's angular momentum that is perpendicular to the planetary orbital plane, $\sqrt{Gm_p a_s(1-e^2)}\cos i$, is conserved, so that the variations in e and i are linked. When the satellite orbital plane is closest to the planetary orbital plane, the eccentricity is maximal. The variations in e and i are thus of opposite algebraic sign for prograde orbits ($i<90°$), but for retrograde orbits ($i>90°$), increasing i brings the orbit closer to the planetary orbital plane so variations in e and i have the same algebraic sign. For $i<i_{crit}=39.2°$ or $i>i_{crit}=140.8°$, the amplitude of the variations in e and i are limited, but for inclinations between the values of i_{crit}, large variations in e can drive the periapse into the region of the regular satellites or the apoapse close to R_H when the argument of periapse ω is circulating. Either consequence can lead to the elimination of the satellite. The only way a high inclination satellite is likely to survive is if ω is librating about 90° or 270° (Kozai resonance) where variations in the eccentricity are more modest. By estimating the fraction of the available phase space for a captured satellite to be in Kozai resonance, Carruba et al. predicted that perhaps 10% of captured satellites should be in stable Kozai resonances at high inclination. The irregular satellites S/2003 J10, Kiviuq (S/2000S5), and Ijiraq (S/2000S6) are apparently currently librating in Kozai resonances. But at least Kiviuq is close to the chaotic zone separating libration from circulation, so its libration may be only temporary (Carruba et al., 2003, 2004).

Numerical integrations of the complete equations of motion of irregular satellites, including the perturbations of the other major planets, show that highly inclined orbits outside the Kozai resonance are more unstable for orbital planes further from the planetary orbital plane between the values of i_{crit} (Carruba et al., 2002). As the extremes of the variable inclinations approach the limits of the zone of exclusion, the satellites can survive for longer periods of time, but the unstable region seems to approach the observed limits asymptotically as the integrations progress. The nonrandom distributions of orbital inclinations and values of the other observed orbital parameters of the irregular satellites are therefore completely consistent with the satellites having been captured into random orbits from the solar nebula population of planetesimals by any of the means discussed in the preceding text.

Appendix A Accretion Disks

The regular satellites of the major planets are in nearly circular orbits in the planes of the planets' equators. This means these satellites had to have formed in dissipative disks of gas and particles that settle to the midplane of the disk before the accretion of the satellites occurs. Given that these satellites

form in such a disk during the formation of the planet, it is important to understand the nature of the accretion in such disks. Our ultimate goal is the determination of the timescales for accretion of such satellites as a function of the disk and particle parameters, for these timescales constrain the current observable properties of the regular satellites. Once formed, the satellites interact with the disks primarily through gas drag when they are small and through density waves generated in the adjacent regions of the disk at mean motion commensurabilities with the satellite when they are large. If the midplane pressure of the disk decreases with distance from the primary, gas drag causes a satellite to spiral toward the planet. Generally the same type of disk causes a similar spiral toward the planet through interaction of the satellite with density waves generated in the disk.

A.1 Accretion of Solids

First, we consider the idealized case of an isolated spherical body m_1 of radius R_1 in a swarm of particles with a distribution $n(m_2, v)\mathrm{d}m_2\,\mathrm{d}v$ representing the number density of particles of mass within $\mathrm{d}m_2$ of m_2 and speeds within $\mathrm{d}v$ of v. If the relative velocities are isotropic, $n(m_2, v)\sin\theta\, d\theta\,\mathrm{d}\phi\,\mathrm{d}m_2\,\mathrm{d}v/4\pi$ represents the number density of the particles with velocities aimed within $\mathrm{d}\theta\,\mathrm{d}\phi$ of spherical polar coordinates θ, ϕ in a coordinate system where the mean particle velocity is zero. One of these particles of radius R_2 and speed v far from m_1 will collide with m_1 during an encounter if the impact parameter is $<R_g$, where

$$R_g^2 = (R_1+R_2)^2\left(1+\frac{v_e^2}{v^2}\right) \qquad [A.1]$$

with $v_e = \sqrt{2G(m_1+m_2)/(R_1+R_2)}$, the escape velocity of m_2 from m_1 when they are in contact. Equation [A.1] is found by conserving energy and angular momentum when m_2 just grazes m_1 during the encounter. The factor $(1+v_e^2/v^2)=F_g$ is called the gravitational enhancement factor for the collisional cross section of m_1 for particles of radius R_2 and speed v relative to m_1. The mass flux of particles with mass within $\mathrm{d}m_2$ of m_2 and velocity within $\mathrm{d}v$ of v directed within $\mathrm{d}\theta\,\mathrm{d}\phi$ of θ, ϕ onto m_1 is $\pi R_g^2 n(m_2,v)m_2 v\,\mathrm{d}m_2\,\mathrm{d}v\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi/4\pi$ such that

$$\dot{m}_1 = \int_{m_2}\int_v \pi n(m_2,v)m_2 v(R_1+R_2)^2\left(1+\frac{2G(m_1+m_2)}{(R_1+R_2)v^2}\right)\mathrm{d}m_2\,\mathrm{d}v \qquad [A.2]$$

is the rate of growth of m_1, if we assume each m_2 that collides is accreted. In eqn [A.2], we have assumed the velocities are isotropic relative to m_1 and integrated over the solid angle and use $R_2=(3m_2/4\pi\rho_p)^{1/3}$, with ρ_p being the mass density of the individual particles. Because $n(m_2, v)$ is not very well constrained, eqn [A.2] is often replaced by the much simpler form

$$\dot{m}_1 \approx \frac{\sigma_s}{2H_s}v\pi(R_1+R_2)^2F_g \qquad [A.3]$$

where R_2, v_e, and v are now suitably averaged values. We have replaced the spatial mass density of the particles by $\sigma_s/2H_s$, where σ_s is the surface mass density of solids and $2H_s$ is the disk thickness. H_s depends on the velocity dispersion v and can be estimated by

$$0.5v_\perp^2 = \int_0^{H_s}(GMz/r^3)\,\mathrm{d}z = 0.5\Omega_K^2H_s^2 \qquad [A.4]$$

where $v_\perp = Kv$ $(K=O(1))$ is the component of the averaged velocity that is perpendicular to the plane of the disk, G is the gravitational constant, M is the mass of the primary, r is the distance from the primary, z is the vertical distance from the midplane, and $\Omega_K=\sqrt{GM/r^3}$ is the Kepler orbital angular velocity of a test particle at distance r from the primary. GMz/r^3 is just the local vertical component of the gravitational acceleration due to the primary. Hence, $H_s = v_\perp/\Omega_K = Kv/\Omega_K$. This value of H_s is generally smaller than the gas pressure scale height c_s/Ω_K (derived later in the text), with c_s being the speed of sound, because of settling of particles toward the midplane.

From eqn [A.3], we can write

$$\dot{R}_1 = \frac{\sigma_s\Omega_K}{8K\rho_p}\left(1+\frac{R_2}{R_1}\right)^2F_g \qquad [A.5]$$

where ρ_p is the mass density of the particles as distinguished from ρ_s, the spatial density of solids in the disk. If we assume $R_2 \ll R_1$ and ignore factors of order unity, we can write the timescale for the growth of a body as (Canup and Ward, 2002)

$$\tau_A = \frac{R_1}{\dot{R}_1} \approx \frac{\rho_p R_1}{\Omega_K\sigma_s F_g} \sim 50\,\text{years}\left(\frac{R_1}{2500\,\text{km}}\right)\left(\frac{\rho_p}{2\,\text{g cm}^{-3}}\right)\left(\frac{10}{F_g}\right)\left(\frac{3\times10^3\,\text{g cm}^{-2}}{\sigma_s}\right)\left(\frac{r}{15R_J}\right)^{3/2} \qquad [A.6]$$

where the numerical form is for the accretion of Ganymede in a minimum mass disk of particles. The surface mass density σ_s is estimated by spreading Ganymede's mass over an annulus extending from the midpoint between Ganymede's and Europa's orbits to the midpoint between Ganymede's and Callisto's orbit to yield $\sigma_s = 3300\,\text{g cm}^{-2}$. With $\rho_p \approx 2\,\text{g cm}^{-3}$ and an orbital radius of $15R_J$, we arrive at $\tau_A \sim 50$ years when $F_g = 10$. This is a lower bound on the timescale for the growth of Ganymede, since σ_s is reduced as the growth proceeds, but it is indicative of the very short timescales for accretion in minimum mass disks.

Safronov (1969), by transforming from relative to random velocities and assuming the vertical and horizontal velocity dispersions are equal, found $K=\pi/4\sqrt{2}$. Safronov also assumed that the velocity dispersion of the accreting particles remained comparable to the escape velocity of the largest embryos as R_1 increased. In this case, $F_g=O(3)$, and R_1 increases linearly with time if the surface mass density σ_s is maintained in the vicinity of m_1. This linear increase in the radius of an accreting body is called orderly growth. $F_g \approx 3$ is not that much smaller than our choice of $F_g=10$ in eqn [A.6]. Runaway growth can make these accretion timescales even shorter.

The assumption by Safronov that the velocity dispersion of the smaller particles was always near the escape velocity of the largest bodies proved to be incorrect in numerical simulations by Wetherill and Stewart (1989). As long as most of the mass of the disk was in the smaller bodies, the velocity dispersion was nearly the escape velocity of the smaller particles. Increases in the velocities of small particles by close encounters with the

larger bodies were damped by interactions with members of the swarm. F_g is thereby increased by orders of magnitude from that assumed in orderly growth and led to what is called runaway growth. For v fixed near the escape velocity of the smaller bodies, the rate of growth in eqn [A.5] becomes proportional to v_e^2, which, for constant density, increases as R_1^2. The change in scaling results in the largest planetesimal growing much faster than any other, and it 'runs away.' For the two largest planetesimals of the swarm of radii $R_1 > R_2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{R_1}{R_2}\right) = \frac{R_1}{R_2}\left(\frac{\dot{R}_1}{R_1} - \frac{\dot{R}_2}{R_2}\right) \approx \frac{R_1}{R_2}(R_1 - R_2) \qquad [A.7]$$

shows that if initially $R_1 > R_2$, the ratio R_1/R_2 grows as long as the mass available for accretion has not decreased substantially.

The timescale for accretion in eqn [A.6] is reduced by the larger value of F_g. Both orderly growth and runaway growth timescales depend on the maintenance of the surface mass density of the disk near the accreting satellite and that most of the disk mass is in the small particles. These timescales increase as the nearby material is exhausted and must be eventually governed by how rapidly particles can be brought into the so-called feeding zone of the growing satellite. Still, numerical simulations of lunar accretion in a particle disk verify the rapid accretion timescales even while accounting for the growth of the average particle size during the process (Ida et al., 1997; Kokubo et al., 2000). Mosqueira and Estrada (2003a,b) determined accretion timescales governed by how rapidly gas drag can bring larger embryos to the feeding zone of the satellite.

Although some stages of the accretion process in disks remain unclear, once small solid objects have formed or are deposited in a flat dissipative medium, their accumulation into a coplanar system of satellites seems reasonably well understood in spite of continued haggling over details. The timing of the formation depends on the temperature history of the disk when ices are the predominant constituents of the forming satellites. Once formed, the satellites will spiral toward the primary because of gas drag and interaction with density waves in the gaseous part (type I drift) (Ward, 1997). It is thus likely that at least the large regular satellites formed during the last stages of accretion of their host planet as the gas in the disk and the solar nebula was being dissipated. The small equatorial satellites residing close to their primaries have most likely been repeatedly disintegrated and reformed in a gas-free environment by planetesimals remaining in the heliocentric swarm. Instabilities in the distribution of large objects lead to mergers and eventual separation of the remaining objects by more than several Hill sphere radii, where the Hill sphere radius (sphere of influence) is defined as the distance from the satellite to its inner Lagrange point ($R_H \approx r(M_s/3M_P)^{1/3}$).

Refinements to the accretion process in disks take account of the effect of the primary and depletion of accreting particles in the feeding zone of a growing object (Greenzweig and Lissauer, 1990, 1992; Lissauer, 1987; Wetherill and Cox, 1984, 1985), where an upper bound on the gravitational enhancement factor of about 1000 is determined. Gas drag replenishes the feeding zone and thereby prolongs the accretion of a body in a disk, but only a fraction of such particles (10–40%) brought into the feeding zone by gas drag are actually accreted (Kary and Lissauer, 1995; Kary et al., 1993).

The previously mentioned discussion is focused on gas-free accretion of solids. However, particles near meter size do not tend to stick to one another during collisions, and they tend to spiral into the planet (or star) because of the head wind created by the less than Kepler speed of orbiting by the gas. Youdin and Goodman (2005) described a streaming instability that can concentrate boulders into quite large gravitationally unstable collections. The streaming instability arises because a concentration of boulders can force the gas within the concentration to orbit closer to the Kepler velocity. The interaction with the surrounding gas streaming by, also containing isolated boulders that are spiraling in, causes the concentration of boulders to increase and become gravitationally bound and collapse to large objects (planetesimals in circumstellar disks) as the relative kinetic energy of the boulders within the swarm is dissipated. In the solar nebula, these objects can exceed the size of the dwarf planet Ceres (Johansen et al., 2007, 2011, 2012). The streaming instability appears to be crucial in the accretion of solids in accretion disks to greater than kilometer size.

A.2 Viscous Spreading

An inherent property of disks orbiting a primary is that they tend to spread as angular momentum is exchanged between adjacent regions because of the Kepler shear. Gravitational, magnetic, and viscous torques and wave transport with shock dissipation are all possible candidates for angular momentum transport (Balbus, 2003). Viscous interchange results from the Kepler shear in the motion of the material orbiting the primary. If we represent the shear stress as due to a viscosity η, we can write the shear stress at a distance r from the primary $\omega_{r\phi} = \eta|\mathrm{d}v_\phi/\mathrm{d}r|_{\mathrm{rot}} = \eta r|\mathrm{d}\Omega_K/\mathrm{d}r| = 3\eta\Omega_K/2$, where $|\mathrm{d}v_\phi/\mathrm{d}r|_{\mathrm{rot}}$ is the radial gradient of the azimuthal velocity in the frame rotating at the Kepler angular velocity at distance r. For a gas,

$$\eta = \frac{m\vec{v}}{3A} \approx \frac{mc_s}{3A} \qquad [A.8]$$

where m is the mean molecular mass, $\vec{v} = \sqrt{8kT/\pi m} \approx c_s$ is the mean molecular speed with k being Boltzmann's constant, T is the temperature, A is the collision cross section of the molecules, and c_s is the sound speed (e.g., Sears and Salinger, 1975). Most of the regular satellites around the major planets have a substantial fraction of ice. So, if ice is to be in solid form as satellitesimals, disk temperatures must be $\lesssim 200$ K (depending on the pressure), so for molecular hydrogen, $c_s \approx 10^5$ cm s^{-1}. The kinematic viscosity $\nu = \eta/\rho_g = 2H_g\eta/\sigma_g \approx c_s H_g m/(A\sigma_g)$, where ρ_g is the gas density, σ_g is the surface mass density of the gas, and H_g is the half thickness of the gas disk.

For the gaseous disk, we choose H_g to be the scale height of the disk that is determined by isostasy in the direction perpendicular to the disk, $\mathrm{d}P/\mathrm{d}z = -\rho_g g_z = -mPg_z/kT = -mPz\Omega_K^2/kT$, where P is the pressure and $g_z = GMz/r^3 = z\Omega_K^2$ is the component of gravity in the z direction. The solution of this equation gives $P = P_0 \exp -z^2/H_g^2$ where $H_g = \sqrt{2kT/m}/\Omega_K \approx c_s/\Omega_K$ is the isothermal scale height. As an example, we can estimate the timescale for viscous transfer of angular momentum in a disk surrounding Jupiter that has the minimum mass for accreting

Ganymede at Ganymede's distance from Jupiter augmented with volatiles to reach solar composition. The amount of solids in the disk necessary to make Ganymede corresponds to a surface density of 3300 g cm^{-2} spread from the midpoint between Ganymede and Europa, and the midpoint between Ganymede and Callisto is thereby increased to $\sim 10^5$ g cm^{-2} such that the kinematic viscosity $\nu \approx 3 \times 10^5$ cm^2 s^{-1}. The timescale for viscous transfer of angular momentum at Ganymede's distance from Jupiter ($\sim 10^{11}$ cm) is $r^2/\nu = 1.2 \times 10^9$ years. This time is so long compared with other timescales associated with a circumplanetary disk that molecular viscosity cannot be important in the disk evolution.

If the gas is turbulent, a turbulent viscosity is invoked to increase the efficiency of angular momentum transport. In this approximation, the molecular mass in eqn [A.8] is replaced by the mass of a typical turbulent eddy $\sim \rho_g l^3$ and the cross section A by $\sim l^2$ where l is a typical eddy size. Finally, $\bar{v}$ is replaced by the mean random velocity of the turbulent eddy v_t, so $\eta_t = \rho_g l v_t$. Next, $v_t = \alpha c_s$, where α is a free parameter, $l \sim H_g = c_s/\Omega_K$, so that the kinematic viscosity appropriate for a turbulent medium is (Shakura and Sunyaev, 1973)

$$\nu_t = \frac{\alpha c_s^2}{\Omega_K} \qquad [A.9]$$

The free parameter α is used to indicate the strength of the non-molecular viscous coupling, and the models are often called 'α disks.' The choices of α are usually based on unsupported assertions about the strength of the turbulence. Lower bounds on the magnitude of angular momentum transport follow from the masses of the giant planets. An equivalent viscosity must be sufficient to prevent the forming planet from opening a gap in the protoplanetary disk before it has acquired a significant fraction of its observed mass (Lin and Papaloizou, 1985). Typically, assumed values of $\alpha = O(10^{-3})$ in applications. For shear stress $w_{r\phi}$, the torque across a cylindrical surface of radius r centered on the primary is $T_\nu = w_{r\phi} 2\pi r^2 2H_g = 6\pi\eta\Omega_K r^2 H_g = 3\pi\alpha\sigma_g c_s^2 r^2$ for an α disk. For the example of a minimum mass disk at Ganymede's distance from Jupiter with $c_s = 10^5$ cm s^{-1} and $\alpha = 10^{-3}$, $\nu_t \approx 10^{12}$ cm^2 s^{-1} and the timescale for viscous spreading $\sim$400 years.

If we consider the mass dm in the cylindrical region of radial extent dr with specific angular momentum h, we have $D(h\mathrm{d}m)/Dt = -(\mathrm{d}T_\nu/\mathrm{d}r)\mathrm{d}r$ to balance the net torque on the cylinder with the change in the angular momentum, where $D/Dt \equiv \partial/\partial t + \mathbf{v}\cdot\nabla$ is the total derivative. In a steady-state disk,

$$\nu_r \frac{\mathrm{d}m}{\mathrm{d}r}\frac{\mathrm{d}h}{\mathrm{d}r} = F_r \frac{\mathrm{d}h}{\mathrm{d}r} = -\frac{\mathrm{d}T_\nu}{\mathrm{d}r} \qquad [A.10]$$

where F_r is the radial mass flux. For the general case where $T_\nu = 0$ at $r = 0$ and as $r \to \infty$ and with $\mathrm{d}h/\mathrm{d}r > 0$ as for Kepler motion, we see that $\mathrm{d}T_\nu/\mathrm{d}r$ must be positive at small r and negative at large r so that mass moves inward for small r and outward for large r. Angular momentum always moves outward.

Particle disks also spread from angular momentum transfer via particle collisions. For a disk of particles of mass m_p and velocity dispersion $\bar{v}$ about the Kepler velocity at distance r from the primary, the equivalent kinematic viscosity is given by $\nu_s = \eta_s/\rho_s = m_p\bar{v}/(3A\rho_s) = 2m_p\bar{v}H_s/(3A\sigma_s) = 4H_s\bar{v}\rho_p R_p/(9\sigma_s)$ (Daisaka et al., 2001), where σ_s is the surface mass density of particles in the disk, R_p is the particle radius, and ρ_p is the particle density (distinguished from the spatial mass density of particles ρ_s). For a disk of 100 m particles corresponding to the minimum mass particle disk at Ganymede's distance from Jupiter and the assumption $\bar{v} = v_{esc}$ of two touching particles, $H_s = Kv_{esc}/\Omega_K$ and $\rho_p = 1$ g cm^{-3}, ν_s is only a little over 80 cm^2 s^{-1}. So, low dispersion particle disks spread very slowly with timescale $r^2/\nu = 4.5 \times 10^{12}$ years. But inside the Roche radius in the lunar case, gravitational instabilities cause patches of higher-density material that are repeatedly sheared apart, thereby yielding enhanced collisional dissipation with an effective viscosity of $\nu \sim (\pi G\sigma_s)^2/\Omega_K^3$ and a resulting timescale $r^2/\nu_t \sim 1$ year (Ward and Cameron, 1978). This rapid spreading has been confirmed by N-body simulations (Takeda and Ida, 2001). But self-heating during this short timescale spreading is so intense, at least for the protolunar disk, that more silicates are vaporized, thereby decreasing the effective viscosity and allowing longer timescales for spreading (Thompson and Stevenson, 1988). So, the long timescales for spreading of particulate disks from the simple molecular viscosity analogy are not likely to prevail (see Salmon et al., 2010, for a thorough treatment of viscous spreading in particle disks).

A.3 Gas Drag

A forming satellite assumes a Kepler orbit about the primary, but the gas in the disk has its central acceleration reduced by a radial pressure gradient and thereby orbits slower than the satellites. The satellite thus experiences a continuous head wind that causes it to spiral toward the primary (Weidenschilling, 1977; Whipple, 1964). The gas motion is governed by the equation

$$\rho_g \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\rho_g \nabla\Phi - \nabla P \qquad [A.11]$$

where Φ is the gravitational potential per unit mass and P is the pressure. In a steady state, cylindrical coordinates $\ddot{r}, \ddot{\phi}, \ddot{z}, \dot{r}$ and $\dot{z}$ are all zero, and the r component of eqn [A.10] is

$$-\rho_g r\dot{\phi}^2 = -\rho_g \frac{\partial\Phi}{\partial r} - \frac{\partial P}{\partial r} \qquad [A.12]$$

Now, $P = \rho_g kT/m = \sigma_g c_s^2/2H_g = \sigma_g c_s \Omega_K/2$ (m = mean molecular mass). Often, the surface mass density and midplane temperatures are represented by power laws (e.g., Weidenschilling, 1977) like $\sigma_g = \sigma_{g0}(r_0/r)^p$ and $T = T_0(r_0/r)^q$, where σ_{g0} and T_0 are values at distance r_0 from the primary. Then, $\partial P/\partial r = -K\sigma_g\Omega_K c_s/r$, where $K = O(1)$ depends on the choices for p and q. With $\Phi = -GM_p/r$, the solution of eqn [A.11] yields the gas angular velocity

$$\dot{\phi}^2 \equiv \Omega^2 = \frac{GM_p}{r^3} - \frac{K\sigma_g\Omega_K c_s}{\rho_g r^2} = \Omega_K^2 - \frac{2Kc_s^2}{r^2} \qquad [A.13]$$

from which

$$\Omega_K - \Omega \approx \frac{Kc_s^2}{r^2\Omega_K} \qquad [A.14]$$

where $\Omega_K + \Omega \approx 2\Omega_K$ has been used.

For circularly orbiting material, the gas drag force is given by

$$F_d = -C_d \pi R^2 \rho_g v_{rel}^2 \qquad [A.15]$$

where C_d is the drag coefficient, which depends on the Reynolds number $\mathcal{R} = 2v_{re1}R\rho_g/\eta$, where η is the molecular viscosity,

and $v_{rel} = r(\Omega_K - \Omega) = Kc_s^2/r\Omega_K$ is the relative velocity between the forming satellite and the gas, that is, the head wind. The Reynolds number dependence of C_d (e.g., Prandtl, 1952) leads to $C_d = 20/\mathcal{R}$ (Stokes drag) for $\mathcal{R} \lesssim 0.5$ and $C_d \approx 0.43$ for $\mathcal{R} \gtrsim 500$ (see Peale (1993) for an application of gas drag with full Reynolds number dependence for C_d). The torque on the satellite is rF_d, which when equated to the time derivative of the angular momentum given by $m_p\sqrt{GM_p r}$ yields an expression for $\dot{r}$. The timescale for gas drag changes in the orbit is then

$$\tau_{GD} \equiv \frac{r}{\dot{r}} = \frac{4}{3K^2}\frac{1}{C_d}\frac{\rho_p R}{\Omega_K \sigma_g}\left(\frac{r\Omega_K}{c_s}\right)^3 \sim 10^3 \text{ years}\left(\frac{10}{C_d}\right)\left(\frac{R}{2500\text{km}}\right)\left(\frac{\rho_p}{2\text{gcm}^{-3}}\right) \times \left(\frac{0.1}{c_s/r\Omega_K}\right)^3\left(\frac{3\times 10^5 \text{gcm}^{-2}}{\sigma_g}\right)\left(\frac{r}{15R_J}\right)^{3/2} \quad [A.16]$$

where we have used $K = 3/4$ for $p = q = 1$, and the numerical form is for Ganymede (Canup and Ward, 2002), where the minimum mass particle disk has been augmented by ~30 times as much gas to yield solar composition.

A.4 Migration from Density Waves

Satellites also interact with disks by creating spiral density waves at the commensurabilities of the mean motions in the disk with that of the satellite (e.g., Ward, 1997). The resulting distribution of mass in the density waves exerts gravitational torques on the satellite that tend to push the satellite away from the site of the density wave. Hence, the density waves generated inside the satellite orbit tend to increase its angular momentum, whereas those outside tend to decrease it. Simultaneously, the ring material receives or gives up angular momentum, which tends to push the ring material away from the satellite. The distribution of the resonances outside the satellite versus those inside leads to the outside density waves dominating, such that the net effect is for the satellite to spiral toward its primary, called type I drift, with a timescale given by (Canup and Ward, 2002; Ward, 1997)

$$\tau_I \equiv \frac{r}{\dot{r}} \approx \frac{1}{C_a\Omega_K}\left(\frac{M_p}{M_s}\right)\left(\frac{M_p}{r^2\sigma_g}\right)\left(\frac{c_s}{r\Omega_K}\right)^2 \sim 10^2 \text{ years}\left(\frac{3}{C_a}\right)\left(\frac{2500\text{km}}{R}\right)^3\left(\frac{2\text{gcm}^{-3}}{\rho_p}\right)\left(\frac{c_s/r\Omega_K}{0.1}\right)^2 \left(\frac{3\times 10^5 \text{gcm}^{-2}}{\sigma_g}\right)\left(\frac{15R_J}{r}\right)^{1/2} \quad [A.17]$$

where C_a is a torque asymmetry parameter that is a function of the disk's radial surface density and temperature profiles (e.g., Tanaka et al., 2002; Ward, 1997); M_p and M_s are the planet and satellite masses, respectively; and r is the distance from the M_p center of mass. The last term in eqn [A.17] is $(H_g/r)^2$. The numerical representation is that centered on Ganymede in a minimum mass Jovian disk. Recent works (e.g., Kley and Crida, 2008; Paardekooper and Mellema, 2006a,b) have developed revised type I torque formula for the case of non-isothermal, adiabatic disks, which depend on local disk properties. These are discussed in the Jupiter section.

The repulsion of the gas disk by the satellite can eventually open a gap in the disk if (e.g., Ward and Hahn, 2000)

$$\frac{M_s}{M_p} > c_v\sqrt{\alpha}\left(\frac{c_s}{r\Omega_K}\right)^{5/2} \sim 10^{-4}c_v\left(\frac{\alpha}{10^{-3}}\right)^{1/2}\left(\frac{c_s/r\Omega_K}{0.1}\right)^{5/2} \quad [A.18]$$

where $c_v = O(1\text{–}10)$. This form of the criterion comes from Ward and Canup (2002), where they pointed out that the limiting mass is just above Ganymede's mass for $c_v = 1$ and $\alpha = 10^{-3}$. If a gap is opened, the type I drift is replaced by type II, whose timescale is now the disk viscous timescale (Ward, 1997)

$$\tau_{II} = \frac{r^2}{\nu_t} = \frac{(r/H_g)^2}{\alpha\Omega_K} \sim 300 \text{ years}\left(\frac{10^{-3}}{\alpha}\right)\left(\frac{0.1}{H/r}\right)^2\left(\frac{r}{15R_J}\right)^{3/2}\left(\frac{M_J}{M_p}\right)^{1/2} \quad [A.19]$$

where again the numerical representation is centered on Ganymede and M_p is normalized by the Jupiter mass M_J. Whether or not a gap is opened by a satellite depends on the value of α and the scale height of the disk at the satellite location. For an alternative gap-opening criterion also depending on these two parameters, see Crida et al. (2012).

A.5 Thermodynamics

The temperature distribution in the disk is of fundamental importance in determining nearly all of the parameters associated with accretion within the disk and whether or not the accreted particles survive migration toward the primary. The growing satellites will be near the midplane of the disk, so the midplane temperature T_c must be known. This in turn is determined by an energy balance equation in steady-state disks, where sources of energy include the gravitational energy deposited by the flux of incoming material, the viscous dissipation within the disk, radiation from the still warm primary, and radiation from the surrounding solar nebula. The sink is radiation from the upper and lower surfaces of the disk at some effective temperature, T_e. If the accretion of the satellites is assumed to take place in a disk with no accretion from the solar nebula, the gravitational source of energy now comes from the redistribution of mass during viscous evolution, and the internal energy of the disk is reduced as it cools.

For optically thick disks, $T_e^4 = 2T_c^4/(1 + 3\tau/2) \approx 4T_c^4/3\tau$, whereas for optically thin disks, $T_e^4 \approx 4\tau T_c^4$ (e.g., Shapiro and Teukolsky, 1983), where $\tau = \int_0^{H_g} \kappa\rho_g dz \approx \overline{\kappa}\sigma_g/2$ is the optical depth from midplane to the surface, with κ being the Rosseland mean opacity ($\text{cm}^2\,\text{g}^{-1}$) (e.g., Schwarzschild, 1958). An example of a surface mass distribution in a steady-state disk from Canup and Ward (2002) (Lynden-Bell and Pringle, 1974) is

$$\sigma_g(r) \approx \frac{4\dot{M}}{15\pi\nu_t}\left[\frac{5}{4} - \sqrt{\frac{r_c}{r_d}} - \frac{1}{4}\left(\frac{r}{r_c}\right)^2\right] \text{ for } r < r_c \approx \frac{4\dot{M}}{15\pi\nu_t}\left[\sqrt{\frac{r_c}{r}} - \sqrt{\frac{r_c}{r_d}}\right] \text{ for } r \geq r_c \quad [A.20]$$

where the kinematic viscosity $\nu_t = \alpha c_s^2/\Omega_K$ is the approximation of Shakura and Sunyaev (1973) for a turbulent medium discussed in the preceding text, r_c is the radius inside of which the mass flux $\dot{M}$ is deposited uniformly, and r_d is the edge of the

disk. Another example is a disk with constant mass flux toward the primary throughout the region where the satellites are forming (e.g., Lynden-Bell and Pringle, 1974):

$$\sigma_g(r) = \frac{\dot{M}}{3\pi\nu_t}\left[1-\sqrt{\frac{R_P}{r}}\right] \qquad [A.21]$$

if the viscous couple vanishes at the primary's surface of radius R_P.

Figure 4 shows the Rosalind mean opacities appropriate for low temperatures in an accretion disk. The opacity due to Bell and Lin (1994) accounts for ice grains at low temperatures and 'metallic' grains at high temperatures. Ice particles dominate the opacity up to temperatures slightly above 160 K with approximately a T^2 power law dependence. Above this temperature, the ice evaporates with a sharp drop in opacity and 'metallic' grains dominate the opacity for larger T. The Pollack et al. (1994) opacities include a distribution of organic particles that maintain a high opacity in the region beyond the ice point. These opacities can be modeled by writing $\kappa = \kappa_0 T^{\xi}$, but different values of the exponent ξ apply in different parts of the disk. There have been no studies of the formation of icy satellites with this accounting of the variations in the disk opacities. An example of the energy balance equation for a steady-state disk at distance r from the primary is

$$\sigma_{SB}T_P^4\left[\frac{2}{3\pi}\left(\frac{R_P}{r}\right)^3+\frac{1}{2}\left(\frac{R_P}{r}\right)^2\frac{H}{r}\left(\frac{d\ln H}{d\ln r}-1\right)\right] + F_{grav}(r)+F_{vis}(r) = 2\sigma_{SB}\left(T_e^4-\frac{\beta_n}{4\pi}T_{neb}^4\right) \qquad [A.22]$$

where the first bracketed term on the left is the radiation deposited per unit area from the warm planet (Ruden and Pollack, 1991), $F_{grav}(r)$ is the energy deposited due the accretion, and $F_{vis}(r)$ is the energy deposited per unit surface area from viscous interactions. The fraction of solid angle subtended by the solar nebula is $\beta_n/4\pi$. The most extreme model dependences are in the terms F_{grav} and F_{vis}. There are no constraints on F_{grav} from simulations, and F_{vis} depends on the unknown value of α. The importance of the opacity κ for the midplane properties cannot be overemphasized.

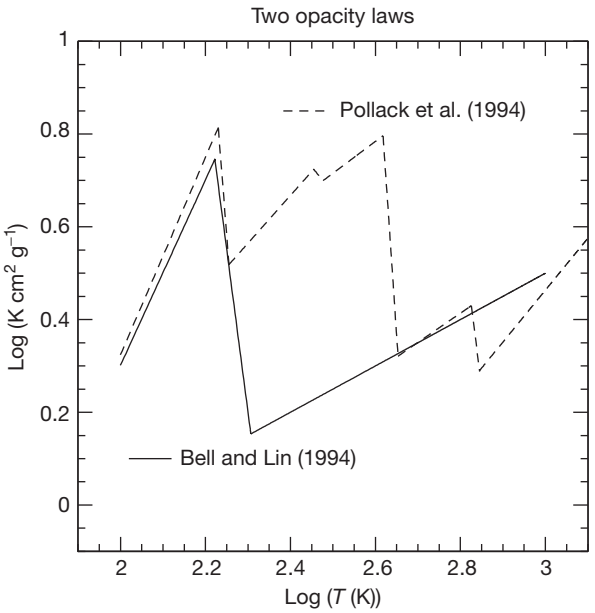


Figure 4 Power law approximations to the Rosseland mean opacities for a cold accretion disk. Adapted from Bell KR and Lin DNC (1994) Using Fu Orionis outbursts to constrain self-regulated protostellar disk models. *Astrophysical Journal* 474: 987–1004; Pollack JB, Hollenbach D, Beckwith S, Simonelli DP, Roush T, and Fong W (1994) Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophysical Journal* 421: 615–639.

Appendix B Tides

Tidal dissipation can drastically change the orbital configuration and the thermal state of the regular natural satellites from the time of their formation. So, it is imperative to understand this evolution, which puts constraints on the conditions of satellite formation. We therefore outline a theory of tidal effects on satellite orbits (and rotations) in this appendix. The potential of mass element dm_0 at $\mathbf{r}$ (within m_0) due to a mass M located at $\mathbf{R}$ relative to a coordinate system fixed in the body m_0 is given by

$$dV = -\frac{GM\,dm_0}{|\mathbf{R}-\mathbf{r}|} = -\frac{GM\,dm_0}{R}\left[1+\frac{r}{R}P_1(\cos S)+\left(\frac{r}{R}\right)^2P_2(\cos S)+\left(\frac{r}{R}\right)^3P_3(\cos S)\ldots\right] \qquad [B.1]$$

where G is the gravitational constant, P_i are the Legendre polynomials, and $\cos S = \mathbf{r}\cdot\mathbf{R}/rR$. The first two terms in the expansion combine to yield an acceleration parallel to $\mathbf{R}$, so the lowest-order tide-raising potential per unit mass in m_0 due to M is given by the second harmonic term

$$W_2 = -\frac{GMr^2}{2R^3}\left[3\cos^2 S-1\right] \qquad [B.2]$$

Typically, $r/R \ll 1$ allows neglect of higher-order terms, but if a satellite is close to a primary (e.g., Phobos), higher-order terms contribute significantly to the total tidal interaction (Bills et al., 2005).

For a homogeneous sphere, Love (1944) showed that for a small distortion, the displacement of the surface due to the disturbing potential W_2 is a small factor ϵ_2 times a surface harmonic S_2 of the same degree:

$$\epsilon_2 S_2 = \frac{5}{2g}\frac{W_2(r=a_e)}{[1+19\mu/(2\rho g a_e)]} \qquad [B.3]$$

where g is the surface gravity acceleration on m_0, μ is the rigidity, ρ is the density, and a_e is the radius of the sphere. The potential per unit mass at $r > a_e$ due to this spherical surface harmonic mass distribution is then

$$-\frac{3}{5}\left(\frac{a_e}{r}\right)^3\epsilon_2 S_2 = -\frac{3}{2[1+19\mu/(\rho g a_e)]}\frac{GM}{2R^3}\times\frac{a_e^5}{r^3}\left(3\cos^2 S-1\right) \qquad [B.4]$$

such that if we place a mass m at $\mathbf{r}$,

$$V_T = -\frac{k_2 GMma_e^5}{2R^5r^5}\left[3(\mathbf{R}\cdot\mathbf{r})^2-r^2R^2\right] \qquad [B.5]$$

is the potential of m at $\mathbf{r}$ due to a tide raised by M on m_0 or the potential of M at $\mathbf{R}$ due a tide raised by m on m_0. The potential Love number $k_2 = 1.5/[1+19\mu/(2\rho g a_e)]$ refers to the body on which the tide is raised. For centrally condensed bodies, k_2 is

reduced. For example, the fluid Love number for the Earth is 0.934 instead of 1.5 (Lambeck, 1980, p. 29).

Now, consider M and m to be the same body. We model the effects of the dissipation by assuming the equilibrium response of m_0 to the tidal stress applied by m lags by some time Δt by analogy to the damped harmonic oscillator forced at a frequency that is small compared to the natural frequency of oscillation. Then, $\mathbf{R}(t)=\mathbf{r}(t-\Delta t)$ is the simplest representation of the effects of dissipation of tidal energy in m_0, that is, $\mathbf{R}$ is the position of m a short time in the past relative to a coordinate system fixed in m_0, where $\mathbf{r}$ is defined in the same coordinate system. The components of $\mathbf{R}=\mathbf{r}(t-\Delta t)$ are not to be included in differentiations of V_T in determining the tidal forces on m. We distinguish these coordinates with a * and write the coordinates of m considered the tide-raising body as

$$\mathbf{r}^*(t-\Delta t)=\mathbf{r}^*(t)-\left.\frac{d\mathbf{r}^*}{dt}\right|_t\Delta t=\mathbf{r}^*+\Delta\mathbf{r}^* \qquad [B.6]$$

If we substitute the right-hand side of eqn [B.6] into eqn [B.5] and expand to the first order in Δr^*, the results were

$$V_T=-k_2Gm^2a_e^5\left(\frac{1}{r^3r^{*3}}+\frac{3\mathbf{r}\cdot\Delta\mathbf{r}^*}{r^4r^{*4}}-\frac{6\mathbf{r}^*\cdot\Delta\mathbf{r}^*}{r^3r^{*5}}\right) \qquad [B.7]$$

where

$$\Delta\mathbf{r}^*=-\left.\frac{d\mathbf{r}^*}{dt}\right|_{body}\Delta t=-\Delta t\left(\frac{dx^{*'}}{dt}\mathbf{e}'_1+\frac{dy^{*'}}{dt}\mathbf{e}'_2+\frac{dz^{*'}}{dt}\mathbf{e}'_3\right) \qquad [B.8]$$

and the time derivative is relative to the $x'y'z'$ system of coordinates fixed in the rotating body m_0. With the generic relation for a vector $\vec{D}$, $\left(d\vec{D}/dt\right)_{body}=\left(d\vec{D}/dt\right)_{space}-\vec{\omega}\times\vec{D}$ and with $\vec{D}\to\mathbf{r}^*$ and $\vec{\omega}\to\vec{\dot{\psi}}$, where $\vec{\dot{\psi}}$ is the angular velocity of m_0, we can write $d\mathbf{r}^*/dt|_{body}=d\mathbf{r}^*/dt|_{space}-\vec{\dot{\psi}}\times\mathbf{r}^*$ and

$$\Delta\mathbf{r}^*\cdot\mathbf{r}^*=-\mathbf{r}\cdot\dot{\mathbf{r}}\Delta t \qquad [B.9]$$

where the derivative on the right-hand side is now relative to inertial space.

As an example, we assume m_0 is rotating about the z' axis at angular velocity $\dot{\psi}$ relative to the xyz axes fixed in inertial space with the xy and $x'y'$ planes being coincident, so that $\psi=\psi_0+\dot{\psi}t$ is the angle between the x and x' axes. Then,

$$\Delta\mathbf{r}^*=-\left(\dot{\psi}y+\dot{x}\right)\Delta t\mathbf{e}_1+\left(\dot{\psi}x-\dot{y}\right)\Delta t\mathbf{e}_2-\dot{z}\Delta t\mathbf{e}_3 \qquad [B.10]$$

The force on m due to the tide it raises on m_0 is found from the negative gradient of V_T in eqn [B.7] with respect to the unstarred coordinates, after which we can set $r^*=r$ and substitute eqns [B.9] and [B.10] into the gradient so obtained to yield

$$\begin{aligned}
-\frac{\partial V_T}{\partial x}&=-k_2Gm^2a_e^5\left[\frac{3x}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}x\Delta t}{r^{10}}+\frac{3\left(\dot{\psi}y+\dot{x}\right)\Delta t}{r^8}\right]\\
-\frac{\partial V_T}{\partial x}&=-k_2Gm^2a_e^5\left[\frac{3y}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}y\Delta t}{r^{10}}+\frac{3\left(-\dot{\psi}x+\dot{y}\right)\Delta t}{r^8}\right]\\
-\frac{\partial V_T}{\partial z}&=-k_2Gm^2a_e^5\left[\frac{3z}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}z\Delta t}{r^{10}}+\frac{3\dot{z}\Delta t}{r^8}\right]
\end{aligned} \qquad [B.11]$$

The gradient $-\nabla V_T$ from eqn [B.11] is the force on m due to tides raised on m_0 by m. If the orbital angular velocity of m is always less than and in nearly the same direction as the spin angular velocity of m_0 (prograde), this force will increase m's orbital semimajor axis and eccentricity and decrease the orbital inclination relative to m_0's equatorial plane. Here, $k_2\to k_2^{m_0}$, the Love number for m_0; $a_e\to R_{m_0}$, the equatorial radius of m_0; $\dot{\psi}\to\dot{\psi}_{m_0}$, the angular velocity of m_0; and $\mathbf{r}\to\mathbf{r}_{m_0m}$, the vector from the center of m_0 to the center of m with corresponding components in inertial space of x, y, and z.

The force on m_0 due to tides raised on m by m_0 has the same form as eqn [B.11] with now $k_2\to k_2^m$, the Love number for m, $m\to m_0$; $a_e\to R_m$, the radius of m; and $\mathbf{r}=\mathbf{r}_{mm_0}=-\mathbf{r}_{m_0m}$. To convert this force to the force on m due to tides raised on m by m_0, the change in sign of the force expressions by Newton's third law is canceled by the change in the sign of the coordinates effected by setting $\mathbf{r}=\mathbf{r}_{\mathbf{m_0m}}$ instead of $-\mathbf{r}_{m_0m}$. So, eqn [B.11] expresses the force on m due to tides raised on m by m_0 with $k_2\to k_2^m$, $m\to m_0$, $a_e\to R_m$ and $\mathbf{r}\to\mathbf{r}_{m_0m}$.

The torque on m due to tides raised by itself on m_0 is $T_T=\mathbf{r}\times(-\nabla V_T)$, and the torque retarding m_0's spin is the negative of this, where $\mathbf{r}=\mathbf{r}_{m_0m}$. Similarly, the torque retarding m's spin due to tides raised on m by m_0 has the same form, but with $\mathbf{r}=\mathbf{r}_{mm_0}$ and $m\to m_0$, $a_e\to R_m$ and $k_2\to k_2^m$. This latter torque on m can be represented by

$$\boldsymbol{T}_T=\frac{3k_2Gm_0^2R^5}{r^6}(\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}_T)(\hat{\mathbf{r}}_T\times\hat{\mathbf{r}}) \qquad [B.12]$$

where we have omitted the sub- and superscripts but remember that the variables apply to m. $\hat{\mathbf{r}}$ is a unit vector toward the tide-raising body m_0, and $\hat{\mathbf{r}}_T$ is a unit vector toward the tidal maximum, which is the subprimary position on m a short time Δt in the past as in the preceding text. Equation [B.12] is valid for arbitrary inclination of the orbit and equatorial planes. Similar to the earlier-mentioned analysis,

$$\hat{\mathbf{r}}_T=\hat{\mathbf{r}}-\frac{d\hat{r}}{dt}\Delta t \qquad [B.13]$$

where the time derivative is relative to the body system of coordinates. Replacement of $\hat{\mathbf{r}}_T$ with eqn [B.13] yields

$$\boldsymbol{T}_T=\frac{3k_2Gm_0^2R^5\Delta t}{r^6}\hat{\mathbf{r}}\times\dot{\hat{\mathbf{r}}} \qquad [B.14]$$

where we have set $\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}_T=1$. So, with the generic relation between the derivatives of a vector relative to body and space coordinates described previously with $\hat{\mathbf{r}}=\cos f\mathbf{e}_1+\sin f\mathbf{e}_2$ and $\boldsymbol{\dot{\psi}}=\dot{\psi}(\sin i\sin\Omega'\mathbf{e}_1-\sin i\cos\Omega'\mathbf{e}_2+\cos i\mathbf{e}_3)+\dot{\Omega}'\mathbf{e}_3$, it is easy to obtain

$$\begin{aligned}
\hat{\mathbf{r}}\times\dot{\hat{\mathbf{r}}}=&\dot{\psi}\sin i\cos(f-\Omega')(-\sin f\mathbf{e}_1+\cos f\mathbf{e}_2)\\
&+\left(\dot{f}-\dot{\Omega}'-\dot{\psi}\cos i\right)\mathbf{e}_3
\end{aligned} \qquad [B.15]$$

where i and Ω' are inclination and longitude of the ascending node of the $x'y'$ (equator) plane on the xy (orbit) plane. Equation [B.15] is substituted into eqn [B.14] and averaged over the orbit period. Useful averages are

$$\begin{aligned}
\left\langle\frac{a^6}{r^6}\dot{f}\right\rangle&=n\left(1+\frac{15}{2}e^2+\frac{45}{8}e^4+\frac{5}{16}e^6\right)/(1-e^2)^6=nf_1(e)\\
\left\langle\frac{a^6}{r^6}\right\rangle&=\left(1+3e^2+\frac{3}{8}e^4\right)/(1-e^2)^{9/2}=f_2(e)\\
\left\langle\frac{a^6}{r^6}\cos 2f\right\rangle&=\left(\frac{3}{2}e^2+\frac{1}{4}e^4\right)/(1-e^2)^{9/2}=f_3(e)\\
\left\langle\frac{a^6}{r^6}\cos^2 f\right\rangle&=\frac{f_2(e)+f_3(e)}{2}\\
\left\langle\frac{a^6}{r^6}\sin^2 f\right\rangle&=\frac{f_2(e)-f_3(e)}{2}
\end{aligned} \qquad [B.16]$$

where $\dot{f} = n\left(a^2\sqrt{1-e^2}\right)/r^2$ has been used in the first of eqn [B.16] with a and e being the orbital semimajor axis and eccentricity, respectively. The averaged tidal torque on m due to tides raised by m_0 on m is thus

$$\langle \boldsymbol{T}_T \rangle = \frac{3k_2 G m_0^2 R^5 \Delta t}{a^6}\left\{-\dot{\psi}\sin i \sin\Omega'\left(\frac{f_2(e)-f_3(e)}{2}\right)\boldsymbol{e}_1 + \dot{\psi}\sin i\cos\Omega'\left(\frac{f_2(e)+f_3(e)}{2}\right)\boldsymbol{e}_2 + \left[nf_1(e) - f_2(e)\dot{\psi}\cos i\right]\boldsymbol{e}_3\right\} \quad [B.17]$$

where we have neglected $\dot{\Omega'}$ compared to n and $\dot{\psi}$.

As an example, consider the retardation of the rotation of a satellite in a circular ($e=0$) and equatorial ($i=0$) orbit from the tidal torque. The time rate of change of the angular momentum of the satellite $C\left(\mathrm{d}\dot{\psi}/\mathrm{d}t\right) = \langle T_T \rangle$, where C is the moment of inertia of m about its spin axis and where only the $\mathbf{e}_3$ term remains in eqn [B.17]. So,

$$\frac{\mathrm{d}\dot{\psi}}{\mathrm{d}t} = -\frac{3k_2 G m_0^2 R^5}{C a^6}\Delta t\left(\dot{\psi}-n\right) = -\frac{15}{2}k_2\frac{m_0}{m}\frac{R^3}{a^3}\frac{n}{Q_0}\left(\dot{\psi}-n\right) \quad [B.18]$$

where $\Delta t = 1/(Q_0 n)$, $C = 2mR^2/5$, and $n^2 = Gm_0/a^3$ have been used. The expression for Δt follows from the discussion in the succeeding text with Q_0 being the dissipation function at frequency n. The solution of eqn [B.18] is a decaying exponential for $\dot{\psi} - n$ (if n is nearly constant) with time constant

$$\tau_T = \frac{2}{15}\frac{1}{k_2}\frac{m}{m_0}\frac{a^3}{R^3}\frac{Q_0}{n} \quad [B.19]$$

where $k_2 = 0.05$ and $Q_0 = 100$, $\tau_T \approx 4300$ years for Ganymede and 57 000 years for Callisto at their current distances from Jupiter. Note that for $e=0$, the asymptotic tidal state is $\dot{\psi} = n$ (i.e., rotation synchronous with the orbital mean motion), but for $e \neq 0$, the asymptotic tidal state is $\dot{\psi} > 0$ determined by the vanishing of $\langle T_T \rangle$ in eqn [B.17]. Satellites with $e \neq 0$ are able to maintain synchronous rotation if the torque on the permanent axial asymmetry dominates $\langle T_T \rangle$ and forces the axis of minimum moment of inertia to librate about the direction to the primary whenever the satellite is at periapse.

The relationship between the dissipation function Q and Δt follows from a simple example. The dissipation parameter Q for a system oscillating at frequency ω is defined by (e.g., Lambeck, 1980, p. 14)

$$\frac{1}{Q} = \frac{\oint \frac{\mathrm{d}E}{\mathrm{d}t}\mathrm{d}t}{2\pi E^*} = \omega\Delta t \quad [B.20]$$

where the numerator is the energy ΔE dissipated during a complete cycle of oscillation and E^* is the maximum energy stored during the cycle. For a tidally distorted, nearly spherical body with the disturbing body in a circular equatorial orbit, a cycle would consist of half a rotation of the distorted body relative to the body causing the tide. For a complex tide generated by a noncircular, nonequatorial orbit of the disturbing body, each periodic term in a Fourier decomposition of the tide would have its own maximum stored energy and dissipation over a complete cycle of oscillation. The response of an oscillator with forcing function $F = A'\sin\omega t$ when $\omega \ll \omega_0$, with ω_0 being the lowest frequency of free oscillation, can be represented by $x = B'\sin\omega(t-\Delta t)$, where Δt is the phase lag in the response due to the dissipation as was assumed in the preceding text for the tidal response. The rate at which the forcing function does work is $\mathrm{d}E/\mathrm{d}t = F\dot{x} = A'B'\omega\sin\omega t\cos\omega(t-\Delta t)$. Then,

$$E(t) = \int_{t_1}^{t} F\dot{x}\,\mathrm{d}t = A'B'\omega\left[\frac{-\cos(2\omega t - \omega\Delta t)}{4\omega} + \frac{\sin\omega\Delta t}{2}t\right]_{t_1}^{t} \quad [B.21]$$

The first term in eqn [B.21] is the periodic storage of energy in the oscillator, and the second is the secular loss of energy. The energy dissipated in a period $2\pi/\omega$ is $\Delta E = A'B'\pi\sin\omega\Delta t$. The peak energy stored is found by integrating $\mathrm{d}E/\mathrm{d}t$ from a value of t where $x=0$ to a value of t where $x = x_{max}$. Since $x \propto \cos(\omega t - \omega\Delta t)$ and $\dot{x} \propto -\omega\sin(\omega t - \omega\Delta t)$, $x=0$ when $\omega t - \omega\Delta t = (2n-1)\pi/2$ and $x = x_{max}$ when $\dot{x} = 0$ or $\omega t - \omega\Delta t = n\pi$. If we choose the limits spanning $t=0$, then the interval of integration is $-\pi/2 + \omega\Delta t \leq \omega t \leq \omega\Delta t$, where $x=0$ at the lower limit and $x = x_{max}$ at the upper limit. Applying these limits to the first term in eqn [B.21] to get the stored energy yields

$$E^* = A'B'\left[\frac{\cos\omega\Delta t}{4} - \frac{-\cos\omega\Delta t}{4}\right] = \frac{A'B'}{2}\cos\omega\Delta t \quad [B.22]$$

Substitution of the expressions for E^* and ΔE into eqn [B.20] yields $1/Q = \tan\omega\Delta t \approx \omega\Delta t$ when the dissipation is relatively small (large Q). Q is then proportional to $1/\omega$ for the assumption in this tidal model that Δt is constant.

The tidal effective Q of the solid Earth has been estimated from satellite observations to be 280 ± 70 (Ray et al., 2001), which is close to the value obtained by the attenuation of normal modes of the Earth (Widmer et al., 1991). For Mars, $Q \approx 100$ from the secular acceleration of the satellite Phobos (Bills et al., 2005; Smith and Born, 1976). For Mars, the dominant frequency is $\dot{\omega} = 2\left|\dot{\psi} - n_P\right|$, where $\dot{\psi}$ is Mars' rotational angular velocity and n_P is Phobos' mean orbital motion. The factor 2 follows from there being two tidal cycles for each synodic period. With a rotation period of 24 h 37 min 23 s and an orbital period of 0.319 days, $\omega = 3.14 \times 10^{-4}$ rad s^{-1} leading to $\Delta t \approx 32$ s for $Q = 100$. For Mercury, the fundamental tidal period is the orbit period, since Mercury rotates 180° relative to the Sun for each orbit. For the attenuation of seismic waves in the Earth and from laboratory experiments on Earth-like materials, Q is proportional to frequency raised to a small, positive fractional (0.2–0.4) power with nominal values near $Q = 100$ (e.g., Shito et al., 2004), but to avoid an infinite discontinuity in the torque whenever a tidal frequency passes through zero while changing sign, an alternative behavior must be invoked (see the discussion on Andrade and Maxwell tidal models).

For energy dissipation within a satellite, the approach of Peale and Cassen (1978) is simplified by Wisdom (2004). The rate at which work is done on a satellite by the tide is given by

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\int_{\mathrm{body}} \rho\mathbf{v}\cdot\nabla W_2\,\mathrm{d}V \quad [B.23]$$

where W_2 is the previously mentioned tide-raising potential and $\mathbf{v}$ is the velocity of a volume element $\mathrm{d}V$ of density ρ. If we

can assume the satellite is incompressible, then $\nabla\cdot\mathbf{v}=0$, and from $\nabla\cdot W_2\mathbf{v}=\mathbf{v}\cdot\nabla W_2+W_2\nabla\cdot\mathbf{v}$, we can write

$$\frac{dE}{dt}=-\int_{body}\rho\nabla\cdot(W_2\mathbf{v})\,dV=-\rho\int_{surface}W_2\mathbf{v}\cdot n\,dS \qquad [B.24]$$

where the last form follows from divergence theorem. In eqn [B.24], $\mathbf{n}$ is the normal to a surface element dS, and we have taken ρ outside the integral by assuming a uniform density.

The rate at which the height of the surface changes is $\mathbf{v}\cdot\mathbf{n}=d\Delta R/dt\approx h_2 dW_2'/dt/g$ (eqn [B.3]), where $h_2=5\,k_2/3$ is the second degree displacement Love number (Love, 1944) and g is the surface gravity acceleration of the satellite. The prime on W_2 indicates the time delay in the response of the body to the forcing function. Then,

$$\frac{dE}{dt}=-\frac{\rho h_2}{g}\int_{surface}W_2\frac{d}{dt}(W_2')\,dS \qquad [B.25]$$

If higher-order terms are necessary in the tidal potential for close satellites, eqn [B.25] would be a sum of terms involving the Love numbers h_3, h_4, etc.

If we consider the case where the satellite equatorial and orbital planes are coincident, $\mathbf{r}=(r\cos f,\ r\sin f,\ 0)$, and $\mathbf{R}=\left(R\cos\left(\phi+\dot\psi t\right),\ R\sin\left(\phi+\dot\psi t\right),\ 0\right)$, where R and ϕ are the ordinary spherical radial and azimuthal coordinates of a point on the surface in a coordinate system fixed in the satellite, and the $\dot\psi t$ accounts for the rotation of the satellite relative to an inertial system. For small eccentricities, we can write $a/r\approx 1+e\cos nt$, $\cos f\approx\cos nt+e(\cos 2nt-1)$ and $\sin f=\sin nt+e\sin 2nt$ to the first order in eccentricity e (e.g., Murray and Dermott, 1999). For the time-lagged potential W_2', t is replaced by $t-\Delta t$ and for a satellite rotating synchronously with its orbital motion, $\dot\psi=n$. With these substitutions, the average rate of energy dissipation is found by performing the surface integral and averaging over an orbit period:

$$\frac{dE}{dt}=-\frac{42}{5}\frac{\pi\rho h_2}{g}\frac{G^2m_p^2R^6}{a^6}e^2n\sin n\Delta t\approx-\frac{42}{19}\frac{\pi\rho^2n^5R^7e^2}{\mu Q} \qquad [B.26]$$

where $h_2\approx 5\rho gR/(19\mu)$ from the definition of k_2, $n^2=Gm_p/a^3$ has been used, and $\sin n\Delta t\to n\Delta t=1/Q$. The last form in eqn [B.26] agrees with that in Peale et al. (1979) as corrected in Peale (2003). Equation [B.26] is the lowest-order approximation for the tidal dissipation in a synchronously rotating satellite with orbital eccentricity e. For very large eccentricities, higher-order terms must be included.

The tidal dissipation in a synchronously rotating satellite will tend to reduce the eccentricity as $e\neq 0$ is the cause of the dissipation. The spin angular momentum of the satellite is conserved because of the synchronous lock. The specific orbital angular momentum $[G(m_p+m_s)a(1-e^2)]^{1/2}$ can thus not gain angular momentum from the satellite spin. The orbital energy $-Gm_pm_s/2a$ must decrease if energy is dissipated in the satellite, and a must thereby decrease. But the conserved angular momentum means e must decrease if a decreases. At the same time, the tide raised on the planet by the satellite tends to increase the eccentricity if the spin angular velocity $\dot\psi_p$ exceeds a value near $1.5n$ for small eccentricity (depending on the tidal model). The rate of change of the eccentricity of a satellite orbit due to tides raised on the primary by the satellite can be determined from the Lagrange planetary equations as expressed in terms of the perturbing accelerations S, T, and W (radial, tangential, and perpendicular to the orbit) (e.g., Plummer, 1960 (1918), p. 151). For the tidal model adopted herein, S, T, and W can be determined from the xyz components given in eqn [B.11]. Substitution of the required S and T components into the Lagrange equation for de/dt averaged over an orbit period yields

$$\frac{de}{dt}=\frac{k_{2p}}{Q_p}\frac{m_s}{m_p}\frac{R_p^5}{a^5}\frac{ne}{(1-e^2)^{13/2}}\left[\frac{\dot\psi_p(1-e^2)^{3/2}}{n}\left(\frac{33}{2}+\frac{99e^2}{4}+\frac{33e^4}{16}\right)-\left(27+\frac{405e^2}{4}+\frac{405e^4}{8}+\frac{135e^6}{64}\right)\right] \qquad [B.27]$$

where Q_p is the value of Q corresponding to frequency of the orbital mean motion n and the subscripts s and p refer to the satellite and planet, respectively. The value of the planetary angular velocity $\dot\psi_p$ above which $de/dt>0$ increases with eccentricity, which follows because the planet has to be rotating sufficiently faster than the orbital motion at periapse to give it the kick that increases the eccentricity. Otherwise, tides raised on the primary will reduce the eccentricity. In the limit of small eccentricity, $\dot\psi_p\gtrsim 1.64n$ for $de/dt>0$. In a model where the tidal potential is separated into its component frequencies with individual phase lags, Goldreich (1963) found that $\dot\psi_p\gtrsim 1.5n$ for positive de/dt.

The magnitude of the rate of decrease in the eccentricity due to tides raised on the satellite can be found in the same way by interchanging the satellite and planetary parameters as discussed in eqn [B.11], where now $\dot\psi_s\equiv n$ leads to $de/dt<0$ for a satellite rotating synchronously with its orbital mean motion for all values of e. The variation in the eccentricity to lowest order in e from tidal dissipation in the planet and the satellite is thus

$$\frac{de}{dt}=\frac{k_{2p}}{Q_p}\frac{m_s}{m_p}\frac{R_p^5}{a^5}e\left(\frac{33}{2}\dot\psi_p-27n\right)-\frac{21}{2}\frac{k_{2s}}{Q_s}\frac{m_p}{m_s}\frac{R_s^5}{a^5}ne \qquad [B.28]$$

where $\dot\psi_p$ and n are the spin angular velocity and orbital mean motion, respectively, and subscripts p and s refer to planet and satellite. (This equation replaces the erroneous eqn [B.21] of this chapter in the first edition of this volume of the *Treatise on Geophysics*.) Whether eccentricity of a satellite orbit increases or decreases as the orbit evolves under tidal evolution depends on the relative magnitude of the two terms in eqn [B.28].

The reader should be aware of the uncertainty and somewhat arbitrariness of tidal models. Our analysis of tidal dissipation using a constant time lag in the response of a planet or satellite to the forcing function leads to $Q\propto 1/\omega$, where ω is a tidal frequency, is the simplest. It has the most straightforward development while yielding relatively intuitive tidal evolution. It has the advantage of being applicable for arbitrary eccentricities, whereas most other models require an expansion in eccentricity (e.g., Kaula, 1964), which must be truncated. Each of the periodic terms in the Kaula expansion is lagged in phase from the same term in the tide-raising potential with the phase lag having the same sign as the frequency of the term. The magnitude of the phase lag is $1/Q$ and one can assume various frequency dependencies for the value of Q to represent different tidal models. Our constant Δt model would have $Q\propto 1/\omega$ where ω is the frequency of the term. A tidal

model often used is Q = constant, which has the nonphysical characteristic that each term in the expansion changes sign abruptly without reducing its magnitude as the frequency of that term passes through zero. Other frequency dependencies for Q can be substituted into the Kaula expansion to represent other assumptions about the dissipative properties of the material.

The major disadvantage of the constant Δt model for solid bodies is that it does not account for creep or residual distortion after the perturbing force is removed. A discussion of how creep is handled is given in Karato's (2008) book, where quite complicated frequency dependencies for Q may result. Some researchers now favor the empirically determined Andrade tidal model (Castillo-Rogez et al., 2011), but it mimics the $Q \propto 1/\omega$ model at very low frequencies and thereby no longer represents real materials. There is some evidence favoring the Maxwell model at very low frequencies, and Efroimsky (2012) had formulated a tidal model with the empirical Andrade model used for normal frequencies but transitioning to the Maxwell model for very low frequencies. All of these manipulations are for solid materials. Although it is usually inferred that partially molten bodies, as in our discussion of the Earth–Moon system, have a high dissipation compared to the completely solid or completely fluid bodies, there is no detailed tidal model to represent this state. The frustration of model dependence of all conclusions regarding the outcome of tidal dissipation persists.

We have seen in the preceding text that nominal values of Q (or Q_0 in the model detailed earlier in the text) are near 100 for Mars and for the damping of seismic waves in the Earth's mantle and near 300 for satellite determinations of the solid Earth response to the Moon and for the damping of seismically excited normal modes of the Earth. $Q = 100$ is an often adopted value representing Earth-like materials. However, the damping of seismic waves in the Moon's upper layers leads to $Q \approx 4000$ at 4 Hz and $Q \approx 7000$ at 8 Hz (Nakamura and Koyama, 1982), but the monthly tidal $Q \approx 37$ and the annual $Q \approx 60$ (Williams et al., 2001). The widely varying values of Q reflect the nature of the materials where the dissipation is taking place, where the dryness of the surface layers of the Moon may lead to the high values of Q, and partial melts in the interior may yield the low tidal Qs. Efroimsky (2012) pointed out the uncertainties in comparing seismic Qs with tidal Qs. Finally, a totally different procedure is needed for the unsolved problem of determining effective Qs of gaseous planets (and stars), where resonances of tidal frequencies with normal modes of oscillation of the gaseous body become important. In summary, the uncertainty of the proper tidal model to use in various circumstances persists, but the simplest tidal theory described previously in the text can be used to determine qualitative constraints on the origin of the natural satellites from their inferred evolutions.

Acknowledgments

We are grateful to David Stevenson for commenting on the Earth–Moon section. The support of SJP by the Planetary Geology and Geophysics Program under grant NNX08AL76G is gratefully acknowledged.

References

Agnor CB, Canup RM, and Levison HF (1999) On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* 100: 307–325.

Agnor CB and Hamilton DP (2004) Satellite capture via binary exchange reactions: Application to Triton. *Bulletin of the American Astronomical Society* 36: 1169.

Agnor CB and Hamilton DP (2006) Neptune's capture of its moon Triton in a binary-planet gravitational encounter. *Nature* 441: 192–194.

Alfven H (1963) The early history of the Moon and the Earth. *Icarus* 1: 357–363.

Alfvén H and Arrhenius G (1969) Two alternatives for the history of the Moon. *Science* 165: 11–17.

Alfvén H and Arrhenius G (1972) Origin and evolution of the Earth-Moon system. *Moon* 5: 210–230.

Alfvén H and Arrhenius G (1976) *Evolution of the Solar System*. Washington, DC: NASA Scientific and Technical Information Office, NASA SP-345.

Alibert Y, Mordasini C, Benz W, and Winisdoerffer C (2005a) Models of giant planet formation with migration and disc evolution. *Astronomy and Astrophysics* 434: 343–353.

Alibert Y and Mousis O (2006) Structure and evolution of Saturn's subnebula – Implications for the formation of Titan. In: *Lunar and Planetary Science XXXVII*, Abstract 1141.

Alibert Y, Mousis O, and Benz W (2005b) Modeling the Jovian subnebula I. Thermodynamic conditions and migration of proto-satellites. *Astronomy and Astrophysics* 439: 1205–1213.

Anderson JD, Jacobson RA, McElrath TP, Moore WB, Schubert G, and Thomas PC (2001) Shape, mean radius, gravity field and interior structure of Callisto. *Icarus* 153: 157–161.

Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998a) Distribution of rock, metals and ice in Callisto. *Science* 280: 1573–1576.

Anderson JD, Scubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998b) Europa's differentiated internal structure: Inferences from four Galileo encounters. *Science* 281: 2019–2022.

Andert TP, Rosenblatt P, Pätzold M, et al. (2010) Precise mass determination and nature of Phobos. *Geophysical Research Letters* 37: L09202, 4 pp.

Andrews-Hanna JC, Asmar SW, Head JW, et al. (2013) Ancient igneous intrusions and early expansion of the Moon revealed by GRAIL gravity gradiometry. *Science* 339: 675–678.

Artymowicz P and Lubow SH (1996) Mass flow through gaps in circumbinary disks. *Astrophysical Journal Letters* 467: L77–L80.

Asphaug E and Reufer A (2013) Late origin of the Saturn system. *Icarus* 223: 544–565.

Balbus S (2003) Enhanced angular momentum transport in accretion disks. *Annual Review of Astronomy and Astrophysics* 41: 555–597.

Banfield D and Murray N (1992) A dynamical history of the inner Neptunian satellites. *Icarus* 99: 390–401.

Barr AC and Canup RM (2008) Constraints on gas giant satellite formation from the interior states of partially differentiated satellites. *Icarus* 198: 163–177.

Barr AC, Citron RI, and Canup RM (2010) Origin of a partially differentiated Titan. *Icarus* 209: 858–862.

Bate MR, Lubow SH, Ogilvie GI, and Miller KA (2003) Three-dimensional calculations of high-and low-mass planets embedded in protoplanetary discs. *Monthly Notices of the Royal Astronomical Society* 341: 213–229.

Belbruno E and Gott JR (2005) Where did the Moon come from? *Astronomical Journal* 129: 1724–1745.

Bell KR and Lin DNC (1994) Using Fu Orionis outbursts to constrain self-regulated protostellar disk models. *Astrophysical Journal* 474: 987–1004.

Benner LAM and McKinnon WB (1995) Orbital behavior of captured satellites: The effect of solar gravity on Triton's postcapture orbit. *Icarus* 114: 1–20.

Benz W, Slattery WL, and Cameron AGW (1986) The origin of the moon and the single-impact hypothesis I. *Icarus* 66: 515–535.

Benz W, Slattery WL, and Cameron AGW (1987) The origin of the moon and the single-impact hypothesis II. *Icarus* 71: 30–45.

Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004.

Borderies N, Goldreich P, and Tremaine S (1984) Excitation of inclinations in ring-satellite systems. *Astrophysical Journal* 284: 429–434.

Borderies N and Yoder CF (1990) Phobos' gravity field and its influenced on its orbit and physical librations. *Astronomy and Astrophysics* 233: 235–251.

Borg LE, Connelly JN, Boyer M, and Carlson RW (2011) Chronological evidence that the Moon is either young or did not have a global magma ocean. *Nature* 477: 70–74.

Boss AP and Peale SJ (1986) Dynamical constraints on the origin of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 59–101. Houston, TX: Lunar and Planetary Institute.

Boué G and Laskar J (2010) A collisionless scenario for Uranus tilting. *Astrophysical Journal* 712: L44–L47.

Bryden G, Chen X, Lin DNC, Nelson RP, and Papaloizou JCB (1999) Tidally induced gap formation in protostellar disks: Gap clearing and suppression of protoplanetary growth. *Astrophysical Journal* 514: 344–367.

Bryden G, R'ozyczka M, Lin DNC, and Bodenheimer P (2000) On the interaction between protoplanets and protostellar disks. *Astrophysical Journal* 540: 1091–1101.

Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2006) Orbits and photometry of Pluto's satellites: Charon, S/2005 P1 and S/2005 P2. *Astronomical Journal* 132: 290–298.

Buie MW, Tholen DJ, and Grundy WM (2012a) The orbit of Charon is circular. *Astronomical Journal* 144, 19 pp.

Buie MW, Tholen DJ, and Grundy WM (2012b) Two body orbits for the satellites of Pluto. In: *Asteroids, Comets, Meteors*, Proceedings of the Conference held in Niigata, Japan, May 16–20. LPI Contribution 1667, id.6249.

Burns JA (1986) Some background on satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 1–38. Tucson, AZ: University of Arizon Press.

Burns JA (2003) Smashed into orbit, a review of *The Big Splat, or How Our Moon Came to Be*, by Dana MacKenzie. *Nature* 425: 664–665.

Cameron AGW and Ward WR (1976) The origin of the Moon (abstract). In: *Lunar Science VII*, pp. 120–122. Houston, TX: Lunar Science Institute.

Canup RM (2004a) Dynamics of lunar formation. *Annual Review of Astronomy and Astrophysics* 42: 441–475.

Canup RM (2004b) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.

Canup RM (2005) A giant impact origin of Pluto–Charon. *Science* 307: 546–550.

Canup RM (2008) Lunar-forming collisions with pre-impact rotation. *Icarus* 196: 518–538.

Canup RM (2010) Origin of Saturn's rings and inner moons by mass removal from a lost Titan-sized satellite. *Nature* 468: 943–946.

Canup RM (2011) On a giant impact origin of Charon, Nix, and Hydra. *Astronomical Journal* 141, 9 pp.

Canup RM (2012) Forming a Moon with an Earth-like composition via a giant impact. *Science* 338: 1052–1055.

Canup RM (2013) Modification of the rock content of the inner Saturnian satellites by an outer solar system LHB. *44th Lunar and Planetary Science Conference*. Contribution No. 1719, p. 2298.

Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.

Canup RM and Esposito LW (1996) Accretion of the Moon from an impact-generated disk. *Icarus* 119: 427–446.

Canup RM and Righter K (2000) *Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.

Canup RM and Ward WR (2000) A possible impact origin of the Uranian satellite system. *Bulletin of the American Astronomical Society* 32: 1105.

Canup R and Ward WR (2001) Steady-state conditions in a Jovian protosatellite accretion disk. Paper presented at the *Jupiter Conference*, Boulder, CO, June 25–30.

Canup R and Ward WR (2002) Formation of the Galilean satellites: Conditions of accretion. *Astronomical Journal* 124: 3404–3423.

Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839.

Canup RM and Ward WR (2009) Origin of Europa and the Galilean satellites. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 59–83. Tucson, AZ: University of Arizona Press.

Canup RM, Levison HF, and Stewart GR (1999) Evolution of a terrestrial multiple-moon system. *Astronomical Journal* 117: 603–620.

Canup RM, Barr AC, and Crawford DA (2013) Lunar-forming impacts: High-resolution SPH and AMR-CTH simulations. *Icarus* 222: 200–219.

Carlson RW (1999) A tenuous carbon dioxide atmosphere on Jupiter's moon Callisto. *Science* 283: 820–821.

Carruba V, Burns JA, Nicholson PD, and Gladman BJ (2002) On the inclination distribution of the Jovian irregular satellites. *Icarus* 158: 434–449.

Carruba V, Nesvorný D, Burns JA, and Ćuk M (2003) Kozai resonators among distant moons of the Jovian planets. *Bulletin of the American Astronomical Society* 35: 1044.

Carruba V, Nesvorný D, Burns JA, Ćuk M, and Tsiganis K (2004) Chaos and the effects of planetary migration on the orbit of S/2000 S5 Kiviuq. *Astronomical Journal* 128: 1899–1915.

Cassen P and Moosman A (1981) On the formation of protostellar disks. *Icarus* 48: 353–376.

Castillo-Rogez JC, Efroimsky M, and Lainey V (2011) The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. *Journal of Geophysical Research* 116: E09008, 29 pp.

Chambers JE (2001) Making more terrestrial planets. *Icarus* 152: 205–224.

Charnoz S, Crida A, Castillo-Rogez JC, et al. (2011) Accretion of Saturn's mid-sized moons during the viscous spreading of young massive rings: Solving the paradox of silicate-poor rings versus silicate-rich moons. *Icarus* 216: 535–550.

Charnoz S, Dones L, Esposito LW, Estrada PR, and Hedman MM (2009) Origin and evolution of Saturn's ring system. In: Dougherty MK, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 537–575. The Netherlands: Springer.

Charnoz S, Salmon J, and Crida A (2010) Origin of Saturn's small moons and F ring: Recent accretion at the ring's outer edge. *Nature* 465: 752–754.

Cheng WH (2011) *Tidal Evolution of Pluto–Charon and the Implication for the Origin of the Satellites Nix and Hydra*. Masters Thesis, University of Hong Kong.

Cheng WH, Lee MH, and Peale SJ (2014a) Complete tidal evolution of Pluto–Charon. *Icarus* 233: 242–258.

Cheng WH, Peale SJ, and Lee MH (2014b) On the origin of Pluto's small satellites by resonant transport. *Icarus* (In press).

Colombo G and Franklin FA (1971) On the formation of the outer satellite groups of Jupiter. *Icarus* 15: 186–189.

Colwell JE and Esposito LW (1992) Origins of the rings of Uranus and Neptune. I – Statistics of satellite disruptions. *Journal of Geophysical Research* 97E: 10227–10241.

Conway BA (1982) On the history of the lunar orbit. *Icarus* 51: 610–622.

Costillo-Rogez JC, Efroimsky M, and Lainey V (2011) The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. *Journal of Geophysical Research* 116: E09008, 29 pp.

Craddock RA (1994) The origin of Phobos and Deimos. In: *Lunar and Planetary Science Conference XXV*, pp. 293–294.

Craddock RA (2011) Are Phobos and Deimos the result of a giant impact? *Icarus* 211: 1150–1161.

Crida A and Charnoz S (2012) Formation of regular satellites from ancient massive rings in the solar system. *Science* 338: 1196–1199.

Ćuk M and Burns JA (2004) Gas-drag-assisted capture of Himalia's family. *Icarus* 167: 369–381.

Ćuk M and Gladman BJ (2005) Constraints on the orbital evolution of Triton. *Astrophysical Journal* 626: L113–L116.

Ćuk M and Stewart ST (2012) Making the Moon from a fast-spinning Earth: A giant impact followed by a resonant despinning. *Science* 338: 1047–1052.

Cuzzi JN, Burns JA, Charnoz S, et al. (2010) An evolving view of Saturn's dynamic rings. *Science* 327: 1470–1475.

Cuzzi JN, Hogan RC, Paque JM, and Dobrovolskis AR (2001) Size selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophysical Journal* 546: 496–508.

Cuzzi JN, Hogan RC, and Shariff K (2008) Toward planetesimals: Dense chondrule clumps in the protoplanetary nebula. *Astrophysical Journal* 687: 1432–1447.

d'Angelo G, Kley W, and Henning T (2003) Orbital migration and mass accretion of protoplanets in three-dimensional global computations with nested grids. *Astrophysical Journal* 586: 540–561.

Daisaka H, Tanaka H, and Ida S (2001) Viscosity in a dense planetary ring with self-gravitating particles. *Icarus* 154: 296–312.

Darwin GH (1879) On the precession of a viscous spheroid, and on the remote history of the Earth. *Philosophical Transactions of the Royal Society* 170: 447–530.

Darwin GH (1880) On the secular changes in the elements of the orbit of a satellite revolving about a tidally distorted planet. *Philosophical Transactions of the Royal Society* 171: 713–891.

Dahl TW and Stevenson DJ (2010) Turbulent mixing of metal and silicate during planet accretion—-And interpretation of the Hf-W chronometer. *Earth and Planetary Science Letters* 295: 177–186.

Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo Program. *Science* 265: 482–490.

Dobrovolskis AR (1982) Internal stresses in Phobos and other triaxial bodies. *Icarus* 52: 136–148.

Dones L (1991) A recent cometary origin for Saturn's rings? *Icarus* 92: 194–203.

Dones L and Tremaine S (1993) On the origin of planetary spins. *Icarus* 103: 67–92.

Duncan MJ, Levison HF, and Lee MH (1998) A multiple time step symplectic algorithm for integrating close encounters. *Astronomical Journal* 116: 2067–2077.

Duxbury TC and Callahan JD (1981) Pole and prime meridian expressions for Phobos and Deimos. *Astronomical Journal* 86: 1722–1727.

Duxbury TC and Callahan JD (1989) Phobos and Deimos control networks. *Icarus* 77: 275–286.

Duxbury TC and Veverka J (1978) Deimos encounter by Viking – Preliminary imaging results. *Science* 201: 812–814.

Dyczmons V (1978) Formation of planetary and satellite systems. *Astrophysics and Space Science* 58: 521–523.

Efroimsky M (2012) Tidal dissipation compared to seismic dissipation in small bodies, Earths and super-Earths. *Astrophysical Journal* 746, 20 pp.

Elkins-Tanton LT, Burgess S, and Yin Q-Z (2011) The lunar magma ocean: Reconciling the solidification process with lunar petrology. *Earth and Planetary Science Letters* 304: 326–336.

Esposito LW (1986) Structure and evolution of Saturn's rings. *Icarus* 67: 345–357.

Estrada PR and Mosqueira I (2006) A gas-poor planetesimal capture model for the formation of giant planet satellite systems. *Icarus* 181: 486–509.

Estrada PR, Mosqueira I, Lissauer JJ, D'Angelo G, and Cruikshank DP (2009) Formation of Jupiter and conditions of accretion of the Galilean satellites. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 27–58. Tucson, AZ: University of Arizona Press.

Farinella P, Milani A, Nobili AM, and Valsecchi GB (1980) Some remarks on the capture of Triton and the origin of Pluto. *Icarus* 44: 810–812.

Fleshman BL, Delamere PA, and Bagenal F (2010) The source of Saturn's extended neutral cloud. *AGU Fall Meeting*.

Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.

Fortney JJ, Marley MS, and Barnes JW (2007) Planetary radii across five orders of magnitude in mass and stellar insolation. *Astrophysical Journal* 659: 1661–1672.

French RS and Showalter MR (2012) Cupid is doomed: An analysis of the stability of the inner Uranian satellites. *Icarus* 220: 911–921.

Fujii YI, Okuzumi S, and Inutsuka S (2011) A fast and accurate calculation scheme for ionization degrees in protoplanetary and circumplanetary disks with charged dust grains. *Astrophysical Journal* 743: 53–62.

Gao P and Stevenson DJ (2012) How does nonhydrostaticity affect the determination of icy satellites' moment of inertia? In: *43rd Lunar and Planetary Science Conference*, 1701.

Garrick-Bethell I, Wisdom J, and Zuber MT (2006) Evidence for a past high eccentricity lunar orbit. *Science* 313: 652–655.

George RB, Halliday AN, Schauble EA, and Reynolds BC (2007) Silicon in the Earth's core. *Nature* 447: 1102–1106.

Gerstenkorn H (1955) Uber Gezeitenreibuing beim Zweikorperproblem. *Zeitschrift für Astrophysik* 36: 245–274.

Gerstenkorn H (1969) The earliest past of the Earth-Moon system. *Icarus* 11: 189–207.

Gladman BJ, Kavelaars JJ, Holman M, et al. (2000) The discovery of Uranus XIX, XX, and XXI. *Icarus* 147: 320–324, erratum (2000) 148, 320.

Gladman B, Kavelaars JJ, Holman M, et al. (2001) Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* 412: 163–169.

Gladman BJ, Nicholson PD, Burns JA, et al. (1998) Discovery of two distant irregular moons of Uranus. *Nature* 392: 897–899.

Goldreich P (1963) On the eccentricity of satellite orbits in the solar system. *Monthly Notices of the Royal Astronomical Society* 126: 257–268.

Goldreich P (1965) Inclination of satellites orbits about an oblate precessing planet. *Astronomical Journal* 70: 5–9.

Goldreich P (1966) History of the lunar orbit. *Reviews of Geophysics* 4: 411–439.

Goldreich P, Lithwick Y, and Sari R (2002) Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* 420: 643–646.

Goldreich P, Murray N, Longaretti PY, and Banfield D (1989) Neptune's story. *Science* 245: 500–504.

Goldreich P and Soter S (1966) Q in the solar system. *Icarus* 5: 375–389.

Goldreich P and Tremaine S (1979) Precession of the epsilon ring of Uranus. *Astronomical Journal* 84: 1638–1641.

Goldreich P and Tremaine S (1982) The dynamics of planetary rings. *Annual Review of Astronomy and Astrophysics* 20: 249–283.

Goldreich P and Ward WR (1973) The formation of planetesimals. *Astrophysical Journal* 183: 1051–1062.

Gomes RS, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Nature* 435: 466–469.

Greenzweig Y and Lissauer JJ (1990) Accretion rates of protoplanets. *Icarus* 87: 40–77.

Greenzweig Y and Lissauer JJ (1992) Accretion rates of protoplanets. II – Gaussian distributions of planetesimal velocities. *Icarus* 100: 440–463.

Halliday AH (2004) Mixing, volatile loss and compositional change during impact driven accretion of the Earth. *Nature* 427: 505–509.

Hamilton DP and Burns JA (1991) Orbital stability zones about asteroids. *Icarus* 92: 118–131.

Hamilton DP and Krivov AV (1997) Dynamics of distant moons of asteroids. *Icarus* 128: 241–249.

Harris AW (1977) The effect of tidal friction on the origin and thermal evolution of the Moon, (abstract). In: *Lunar Science VIII*, pp. 401–402. Houston, TX: Lunar Science Institute.

Harris AW (1978) Satellite formation II. *Icarus* 34: 128–145.

Harris AW and Kaula WM (1975) A co-accretion model of satellite formation. *Icarus* 24: 516–524.

Hartmann WK and Davis DR (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24: 504–515.

Hartmann WK, Phillips RJ, and Taylor GJ (1986) *Origin of the Moon*. Houston, TX: Lunar and Planetary Institute.

Heppenheimer TA and Porco CC (1977) New contributions to the problem of capture. *Icarus* 30: 385–401.

Hibbitts CA, Klemaszewski JE, McCord TB, Hansen GB, and Greeley R (2002) CO_2-rich impact craters on Callisto. *Journal of Geophysical Research (Planets)* 107: 14.1–14.12.

Hibbitts CA, McCord TB, and Hansen GB (2000) Distributions of CO_2 and SO_2 on the surface of Callisto. *Journal of Geophysical Research* 430: 22541–22558.

Hibbitts CA, Pappalardo RT, Hansen GB, and McCord TB (2003) Carbon dioxide on Ganymede. *Journal of Geophysical Research (Planets)* 108: 2.1–2.22.

Hollenbach DJ, York HW, and Johnstone D (2000) Disk dispersal around young stars. In: Manning V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, p. 401. Tucson, AZ: University of Arizona Press.

Holman MJ, Kavelaars JJ, Grav T, et al. (2004) Discovery of five irregular moons of Neptune. *Nature* 430: 865–867.

Horedt GP (1976) Pre-capture orbits of the Moon. *Moon* 15: 439–443.

Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171: 391–410.

Ida S, Canup RM, and Stewart GR (1997) Lunar accretion from an impact generated disk. *Nature* 389: 353–357.

Ida S and Lin DNC (2004) Toward a deterministic model of planetary formation. I. A desert in the mass and semimajor axis distributions of extrasolar planets. *Astrophysical Journal* 604: 388–413.

Ida S and Lin DNC (2008) Toward a deterministic model of planetary formation. IV. Effects of Type I migration. *Astrophysical Journal* 673: 487–501.

Iess L, Rappaport NJ, Jacobson RA, et al. (2010) Gravity science at Titan, *EGU General Assembly,* Vienna, Austria, 2-7 May, P 11037.

Imaeda Y and Inutsuka S (2002) Shear flows in smoothed particle hydrodynamics. *Astrophysical Journal* 569: 501–518.

Jacobsen SB (2005) The Hf-W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences* 33: 531–570.

Jacobson RA (2010) The orbits and masses of the Martian satellites and the libration of Phobos. *Astronomical Journal* 139: 668–679.

Jaumann R, Clark RN, Nimmo F, et al. (2009) Icy satellites: Geological evolution and surface processes. In: Dougherty MK, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 637–681. Dordrecht: Springer.

Jewitt D and Sheppard S (2005) Irregular satellites in the context of planet formation. *Space Science Reviews* 116: 441–455.

Johansen A, Klahr H, and Henning T (2006) Gravoturbulent formation of planetesimals. *Astrophysical Journal* 636: 1121–1134.

Johansen A, Klahr H, and Henning T (2011) High resolution simulations of planetesimal formation in turbulent protoplanetary discs. *Astronomy and Astrophysics* 529, A62, 16 pp.

Johansen A, Oishi J, Mac Low M-M, Klahr H, Henning T, and Youdin A (2007) Rapid planetesimal formation in turbulent circumstellar disks. *Nature* 448: 1022–1025.

Johansen A and Youdin A (2007) Protoplanetary disk turbulence driven by the streaming instability: Nonlinear saturation and particle concentration. *Astrophysical Journal* 662: 627–641.

Johansen A, Youdin AN, and Lithwick Y (2012) Adding particle collisions to the formation of asteroids and Kuiper belt objects via streaming instabilities. *Astronomy and Astrophysics* 537, A125, 17 pp.

Johansen A, Youdin A, and Mac Low M-M (2009) Particle clumping and planetesimal formation depend strongly on metallicity. *Astrophysical Journal* 704: L75–L79.

Jones JH and Palme H (2000) Geochemical constraints on the origin of the Moon. In: Canup RM and Richter J (eds.) *Origin of the Earth and Moon*, pp. 197–216. Tucson, AZ: University of Arizona Press.

Karato S-I (2008) *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth*. Cambridge: Cambridge University Press.

Kary DM and Lissauer JJ (1995) Nebular gas drag and planetary accretion. II. Planet on an eccentric orbit. *Icarus* 117: 1–24.

Kary DM, Lissauer JJ, and Greenzweig Y (1993) Nebular gas drag and planetary accretion. *Icarus* 106: 288–307.

Kaula WM (1964) Tidal dissipation by tidal friction and the resulting orbital evolution. *Reviews of Geophysics* 2: 661–685.

Kivelson MG, Khurana KK, Coroniti FV, et al. (1997) Magnetic field and magnetosphere of Ganymede. *Geophysical Research Letters* 24: 2155–2158.

Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104: 4609–4626.
Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
Kleine T, Mezger KK, Palme H, and Münker C (2004) The W isotope evolution of the bulk silicate Earth: Constraints on the timing and mechanisms of core formation and accretion. *Earth and Planetary Science Letters* 228: 109–123.
Kleine T, Palme H, Mezger K, and Halliday AN (2005) Hf-W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* 310: 1671–1674.
Kley W (1999) Mass flow and accretion through gaps in accretion discs. *Monthly Notices of the Royal Astronomical Society* 303: 696–710.
Kley W and Crida A (2008) Migration of protoplanets in radiative discs. *Astronomy and Astrophysics* 487: L9–L12.
Kley W, D'Angelo G, and Henning T (2001) Three-dimensional simulations of a planet embedded in a protoplanetary disk. *Astrophysical Journal* 547: 457–464.
Kokubo E, Ida S, and Makino J (2000) Evolution of a circumterrestrial disk and formation of a single Moon. *Icarus* 148: 419–436.
Kolyuka YF, Efimov AE, Kudryavtsev SM, Margorin OK, Tarasov VP, and Tikhonov VF (1990) Refinement of the gravitational constant of Phobos from Phobos-2 tracking data. *Soviet Astronomy Letters* 16: 168–170.
König S, Münker C, Hohl S, et al. (2011) The Earth's tungsten budget during mantle melting and crust formation. *Geochimica et Cosmochimica Acta* 75: 2119–2136.
Kuskov OL and Kronrod VA (2005) Internal structure of Europa and Callisto. *Icarus* 177: 550–569.
Lainey V, Arlot J-E, Karatekin Ö, and van Hoolst T (2009) Strong tidal dissipation in Io and Jupiter from astrometric observations. *Nature* 459: 957–959.
Lainey V, Karatekin Ö, Desmars J, et al. (2012) Strong tidal dissipation in Saturn and constraints on Enceladus' thermal state from astrometry. *Astrophysical Journal* 752: 14L.
Lambeck K (1980) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. New York: Cambridge University Press.
Lambrechts M and Johansen A (2012) Rapid growth of gas-giant cores by pebble accretion. *Astronomy and Astrophysics* 544, A32, 13 pp.
Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.
Lee MH and Peale SJ (2000) Making Hyperion. *Bulletin of the American Astronomical Society* 32: 1078.
Lee MH and Peale SJ (2006) On the orbits and masses of the satellites of the Pluto–Charon system. *Icarus* 184: 573–583.
Lin DCN and Papaloizou J (1985) On the dynamical origin of the solar system. In: Black D and Matthews MS (eds.) *Protostars and Planets II*, pp. 981–1072. Tucson, AZ: University of Arizona Press.
Lissauer JJ (1987) Timescales for planetary accretion and structure of the protoplanetary disk. *Icarus* 69: 249–265.
Lissauer JJ (1995) Urey Prize lecture: On the diversity of plausible planetary systems. *Icarus* 114: 217–236.
Lissauer JJ and Kary DM (1991) The origin of the systematic component of planetary rotation: 1. Planet on a circular orbit. Icarus 94: 126–159.
Lithwick Y and Wu Y (2008) On the origin of Pluto's minor moons Nix and Hydra. Astrophysics ArXiv:0802.2951v1.
Love AEH (1944) *A Treatise on the Mathematical Theory of Elasticity*. New York: Dover.
Lubow SH and Martin RG (2013) Dead zones in circumplanetary disks as formation sites for regular satellites. *Monthly Notices of the Royal Astronomical Society* 428: 2668–2673.
Lubow SH, Seibert M, and Artymowicz P (1999) Disk accretion onto high-mass planets. *Astrophysical Journal* 526: 1001–1012.
Lugmair GW and Shukolyukov A (1998) Early solar system timescales according to ^{53}Mn-^{53}Cr systematics. *Geochimica et Cosmochimica Acta* 62: 2863–2886.
Lunine JI and Nolan MC (1992) A massive early atmosphere on Triton. *Icarus* 100: 221–234.
Lunine J and Stevenson DJ (1982) Formation of the Galilean satellites in a gaseous nebula. *Icarus* 52: 14–39.
Lynden-Bell D and Pringle JE (1974) The evolution of viscous disks and the origin of the nebular variables. *Monthly Notices of the Royal Astronomical Society* 168: 603–637.
Lyttleton RA (1967) Dynamical capture of the Moon by the Earth. *Proceedings of the Royal Society of London* A296: 285–292.
Machida MN (2009) Thermal effects of circumplanetary disc formation around proto-gas giant planets. *Monthly Notices of the Royal Astronomical Society* 392: 514–524.
Machida MN, Kokubo E, Inutsuka S-I, and Matsumoto T (2008) Angular momentum accretion onto a gas giant plant. *Astrophysical Journal* 685: 1220–1236.
Machida MN, Kokubo E, Inutsuka S-I, and Matsumoto T (2010) Gas accretion onto a protoplanet and formation of a gas giant planet. *Monthly Notices of the Royal Astronomical Society* 405: 1227–1243.
MacKenzie D (2003) *The Big Splat, or How Our Moon Came to Be*. Hoboken, NJ: Wiley.
Margot J-L, Nolan MC, Benner LAM, et al. (2002) Binary asteroids in the near-Earth object population. *Science* 296: 1445–1448.
Marley MS, Fortney JJ, Hubickyj O, Bodenheimer P, and Lissauer JJ (2007) On the luminosity of young Jupiters. *Astrophysical Journal* 655: 541–549.
Martin RG and Lubow SH (2011) Tidal truncation of circumplanetary disks. *Monthly Notices of the Royal Astronomical Society* 413: 1447–1461.
Massett FS, Morbidelli A, Crida A, and Ferreira J (2006) Disk surface density transitions as protoplanet traps. *Astrophysical Journal* 642: 478–487.
McCord TB, Hansen GB, Clark RN, et al. (1998) Non-water-ice constituents in the surface material of the icy Galilean satellites from the Galileo near-infrared mapping spectrometer investigation. *Journal of Geophysical Research* 103: 8603–8626.
McKinnon WB (1984) On the origin of Triton and Pluto. *Nature* 311: 355–358.
McKinnon WB (1989) On the origin of the Pluto–Charon binary. *Astrophysical Journal* 344: L41–L44.
McKinnon WB (1997) Mystery of Callisto: Is it undifferentiated? *Icarus* 130: 540–543.
McKinnon WB (2006) On convection in ice I shells of outer Solar System bodies, with detailed application to Callisto. *Icarus* 183: 435–450.
McKinnon WB and Leith AC (1995) Gas drag and the orbital evolution of a captured Triton. *Icarus* 118: 392–413.
McKinnon WB, Lunine JI, and Banfield D (1995) Origin and evolution of Triton. In: Cruikshank DP (ed.) *Neptune and Pluto*, pp. 807–877. Tucson, AZ: University of Arizona Press.
Melosh HJ (2007) A hydrocode equation of state for SiO_2. *Meteoritics & Planetary Science* 42: 2079–2098.
Melosh HJ (2009) An isotopic crisis for the giant impact origin of the Moon? *Meteoritics and Planetary Science Supplement*. 72nd Meteoritical Society Meeting, July 13-18, 2009 in Nancy, France; p 5104.
Merline WJ, Weidenschilling SJ, Durda DD, Margot J-L, Pravec P, and Storrs AD (2002) Asteroids do have satellites. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 289–312. Tucson, AZ: University of Arizona Press.
Mignard F (1979) The evolution of the lunar or orbit revisited, I. *Moon and Planets* 20: 301–315.
Mignard F (1981a) The lunar orbit revisited, III. *Moon and Planets* 24: 189–207.
Mignard F (1981b) Evolution of the Martian satellites. *Monthly Notices of the Royal Astronomical Society* 194: 365–379.
Mitchell TR and Stewart GR (2011) Photoevaporation as a truncation mechanism for circumplanetary disks. *Astronomical Journal* 142: 168–178.
Morbidelli A and Crida A (2007) The dynamics of Jupiter and Saturn in the gaseous protoplanetary disk. *Icarus* 191: 158–171.
Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.
Morbidelli A, Tsiganis K, Batygin K, Crida A, and Gomes R (2012) Explaining why the Uranian satellites have equatorial prograde orbit despite the large planetary obliquity. *Icarus* 219: 737–740.
Mosqueira I and Estrada PR (2003a) Formation of the regular satellites of giant planets in an extended gaseous nebula I: Subnebula model and accretion of satellites. *Icarus* 163: 198–231.
Mosqueira I and Estrada PR (2003b) Formation of the regular satellites of giant planets in an extended gaseous nebula II: Satellite migration and survival. *Icarus* 163: 232–255.
Mosqueira I and Estrada PR (2011) On the origin of the Saturnian Moon-Ring system, *42nd Lunar and Planetary Science Conference*. Contribution No. 1608, p2151.
Mosqueira I, Estrada P, and Turrini D (2010) Planetesimals and satellitesimals: Formation of the satellite systems. *Space Science Reviews* 153: 431–446.
Mousis O and Alibert Y (2006) Modeling the Jovian subnebula: II. Composition of regular satellite ices. *Astronomy and Astrophysics* 448: 771–778.
Mousis O, Pargamin J, Grasset O, and Sotin C (2002) Experiments in the NH_3-H_2O system in the [0, 1 GPa] pressure range – Implications for the deep liquid layer of large icy satellites. *Geophysical Research Letters* 29(45): 1–4.
Murray CD and Dermott SF (1999) *Solar System Dynamics*. Cambridge: Cambridge University Press.
Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
Nakajima M and Stevenson DJ (2013) Thermodynamic processes during the Moon-forming impact. In: *44th Lunar and Planetary Conference*, 2680.pdf.
Nakamura Y and Koyama J (1982) Seismic Q of the lunar upper mantle. *Journal of Geophysical Research* 87: 4855–4861.

Nesvorny D, Vokrouhlicky D, and Mobidelli A (2007) Capture of irregular satellites during planetary encounters. *Bulletin of the American Astronomical Society* 39: 475.

Nesvorny D, Youdin AN, and Richardson DC (2010) Formation of Kuiper Belt binaries by gravitational collapse. *Astronomical Journal* 140: 785–793.

Niemann HB, Atreya SK, Bauer SJ, et al. (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438: 779–784.

Niemann HB, Atreya SK, Demick JE, et al. (2010) Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini–Huygens probe gas chromatograph mass spectrometer experiment. *Journal of Geophysical Research* 115: E12006.

Nogueira E, Brasser R, and Gomes R (2011) Reassessing the origin of Triton. *Icarus* 214: 113–130.

O'Rourke JG and Stevenson DJ (2013) Stability of ice/rock mixtures with application to a partially differentiated Titan. *Icarus* 227: 67–77.

Ogihara M, Duncan MJ, and Ida S (2010) Eccentricity trap: Trapping of resonantly interacting planets near the disk inner edge. *Astrophysical Journal* 721: 1184–1192.

Ogihara M and Ida S (2012) N-body simulations of satellite formation around giant planets: Origin of orbital configuration of the Galilean moons. *Astrophysical Journal* 753: 60–77.

Öpik EJ (1969) The Moon's surface. *Annual Review of Astronomy and Astrophysics* 7: 473–526.

Öpik EJ (1971) Comments on lunar origin. *Irish Astronomical Journal* 10: 190–238.

Owen T and Niemann HB (2009) The origin of Titan's atmosphere: Some recent advances. *Philosophical Transactions of the Royal Society A* 367: 607–615.

Paardekooper S-J, Baruteau C, Crida A, and Kley W (2010) A torque formula for non-isothermal type I planetary migration —I. Unsaturated horseshoe drag. *Monthly Notices of the Royal Astronomical Society* 401: 1950–1964.

Paardekooper S-J and Mellema G (2006a) Dust flow in gas disks in the presence of embedded planets. *Astronomy and Astrophysics* 453: 1129–1140.

Paardekooper S-J and Mellema G (2006b) Halting type I planet migration in non-isothermal disks. *Astronomy and Astrophysics* 459: L17–L20.

Pahlevan K and Stevenson DJ (2007) Equilibrium in the aftermath of the lunar-forming impact. *Earth and Planetary Science Letters* 262: 438–449.

Pang KD, Pollack JB, Veverka J, Lane AL, and Ajello JM (1978) The composition of Phobos – Evidence for carbonaceous chondrite surface from spectral analysis. *Science* 199: 64–66.

Papaloizou JCB and Larwood JD (2000) On the orbital evolution and growth of protoplanets embedded in a gaseous disc. *Monthly Notices of the Royal Astronomical Society* 315: 823–833.

Papaloizou JCB and Nelson RP (2005) Models of accreting gas giant protoplanets in protostellar disks. *Astronomy and Astrophysics* 433: 247–265.

Papaloizou JCB and Terquem C (1999) Critical protoplanetary core masses in protoplanetary disks and the formation of short-period giant planets. *Astrophysical Journal* 521: 823–838.

Peale SJ (1977) Rotation histories of the natural satellites. In: *Planetary Satellites*, pp. 87–111. Tucson, AZ: University of Arizona Press.

Peale SJ (1986) Orbital resonances, unusual configurations and exotic rotation states among planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 159–223. Tucson, AZ: University of Arizona Press.

Peale SJ (1988) Speculative histories of the Uranian satellite system. *Icarus* 74: 153–171.

Peale SJ (1993) The effect of the nebula on the Trojan precursors. *Icarus* 106: 308–322.

Peale SJ (1999) Origin and evolution of the natural satellites. *Annual Review of Astronomy and Astrophysics* 37: 533–602.

Peale SJ (2003) Tidally induced volcanism. *Celestial Mechanics and Dynamical Astronomy* 87: 129–155.

Peale SJ and Cassen PM (1978) Contribution of tidal heating to lunar thermal history. *Icarus* 36: 245–269.

Peale SJ, Cassen PM, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.

Peale SJ and Lee MH (2002) A primordial origin of the Laplace relation among the Galilean satellites. *Science* 298: 593–597.

Pires dos Santos PM, Giuliatti-Winter SM, and Sfair R (2011) Gravitational effects of Nix and Hydra in the external region of the Pluto–Charon system. *Monthly Notices of the Royal Astronomical Society* 410: 273–279.

Pires dos Santos PM, Morbidelli A, and Nesvorny D (2012) Dynamical capture in the Pluto–Charon system. *Celestial Mechanics and Dynamical Astronomy* 114: 341–352.

Plummer HC (1960) *Introductory Treatise on Dynamical Astronomy*. New York: Dover, Reprinted from the Cambridge University Press edition (1918).

Pollack JB and Burns JA (1977) An origin by capture for the Martian satellites? *Bulletin of the American Astronomical Society* 9: 518.

Pollack JB, Burns JA, and Tauber ME (1979) Tauber gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* 37: 587–611.

Pollack JB, Hollenbach D, Beckwith S, Simonelli DP, Roush T, and Fong W (1994) Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophysical Journal* 421: 615–639.

Pollack JB, Veverka J, Pang KD, Colburn DS, Lane AL, and Ajello JM (1978) Multicolor observations of Phobos with the Viking lander cameras – Evidence for a carbonaceous chondritic composition. *Science* 199: 66–69.

Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.

Prandtl L (1952) *Essentials of Fluid Dynamics*. New York: Hafner, pp. 174–196.

Pritchard ME and Stevenson DJ (2000) Thermal aspects of a lunar origin by giant impact. In: Canup RM and Richter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tucson, AZ: University of Arizon Press.

Ray R, Eanes RJ, and Lemoine FG (2001) Constraints on energy dissipation in the Earth's body tide from satellite tracking and altimetry. *Geophysical Journal International* 144: 471–480.

Reufer A, Meier MMM, Benz W, and Wieler R (2012) A hit and run giant impact scenario. *Icarus* 221: 296–299.

Ringwood AE (1970) Origin of the Moon: The precipitation hypothesis. *Earth and Planetary Science Letters* 8: 131–140.

Ringwood AE (1972) Some comparative aspects of lunar origin. *Physics of the Earth and Planetary Interiors* 6: 366–376.

Rosenblatt P (2011) The origin of the Martian moons revisited. *Astronomy and Astrophysics* 19: 44.

Rosenblatt P and Charnoz S (2012) On the formation of the Martian moons from a circum-Mars accretion disk. *Icarus* 221: 806–815.

Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal-silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

Ruden SP and Pollack JB (1991) The dynamical evolution of the protosolar nebula. *Astrophysical Journal* 375: 740–760.

Ruskol EL (1961) The origin of the Moon I. *Soviet Astronomy-AJ* 4: 657–668.

Ruskol EL (1963) The origin of the Moon II. *Soviet Astronomy-AJ* 7: 221–227.

Ruskol EL (1972a) The origin of the Moon. III. Some aspects of the dynamics of the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 646–654.

Ruskol EL (1972b) Possible differences in the chemical composition of the Earth and Moon, for a Moon formed in the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 1061–1063.

Ruskol EL (1972c) On the initial distance of the Moon forming in the circumterrestrial swarm. *Moon* 5: 402–404.

Ruskol EL (1975) The origin of the Moon. In: *Proceedings of the Soviet-American Conference on Cosmo-chemistry of the Moon and Planets*, pp. 638–644. Moscow: Nauka (translated in English in NASA SP-370 (1977), pp. 815–822).

Safronov VS (1969) *Evolution of the Protoplanetary Cloud and the Formation of the Earth and Planets*. Moscow: Nauka Press, NASA TTF-677, 1972.

Safronov VS, Pechernikova GV, Ruskol EL, and Vitiazev AV (1986) Protosatellite swarms. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 89–116. Tucson, AZ: University of Arizona Press.

Saha P and Tremaine S (1993) The orbits of the retrograde Jovian satellites. *Icarus* 106: 549–562.

Salmon J and Canup RM (2012) Lunar accretion from a Roche-interior fluid disk. *Astrophysical Journal* 760: 18.

Salmon J, Charnoz S, Crida A, and Brahic A (2010) Long-term and large-scale viscous evolution of dense planetary rings. *Icarus* 209: 771–785.

Sasaki T, Stewart GR, and Ida S (2010) Origin of the different architectures of the Jovian and Saturnian satellite systems. *Astrophysical Journal* 714: 1052–1064.

Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories compositions and internal structures of the moons in the solar system. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 224–292. Tucson, AZ: University of Arizona Press.

Schwarzschild M (1958) *Structure and Evolution of the Stars*. Princeton, NJ: Princeton University Press, Dover Publications, New York (1965), p. 62.

Sears FW and Salinger GL (1975) *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Reading, MA: Addison-Wesley, p. 286.

Sekine Y and Genda H (2012) Giant impacts in the Saturnian system: A possible origin of diversity in the inner mid-sized satellites. *Planetary and Space Science* 63–64: 133–138.

Shakura NI and Sunyaev RA (1973) Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics* 24: 337–355.

Shapiro SL and Teukolsky SA (1983) *Black Holes, White Dwarfs, and Neutron Stars*. New York: Wiley, ch. 14.

Shenk P and Jackson MP (1993) Diapirism on Triton – A record of crustal layering and instability. *Geology* 21: 299–302.

Shito A, Karato S, and Park J (2004) Frequency dependence of Q in Earth's upper mantle inferred from continuous spectra of body waves. *Geophysical Research Letters* 31: L12603.
Shock EL and McKinnon WB (1993) Hydrothermal processing of cometary volatiles – Applications to Triton. *Icarus* 106: 464–477.
Shor VA (1975) The motion of the Martian satellites. *Celestial Mechanics* 12: 61–75.
Showalter MR, Hamilton DP, Stern SA, Weaver HA, Steffl AJ, and Young LA (2011) New satellite of (134340) Pluto: S/2011 (134340) 1, IAU Circulars, 9221, 1 (2011), Green DWE (ed.).
Showalter MR, Weaver HA, Stern SA, et al. (2012) New Satellite of (134340) Pluto S/2012 (134340)1. *International Astronomical Union*. Circular 9253, 1.
Shu F (1984) Waves in planetary rings. In: Greenberg R and Brahic A (eds.) *Planetary Rings*, pp. 513–561. Tucson, AZ: University of Arizona Press.
Shu F, Johnston D, and Hollenbach D (1993) Photoevaporation of the solar nebula and the formation of the giant planets. *Icarus* 106: 92–101.
Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure, and dynamics of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, vol. 1. New York: Cambridge University Press.
Shultz PH and Lutz-Garihan AB (1982) Forgotten satellites of Mars: A possible record from oblique angle impacts. In: *Lunar and Planetary Science XIII*, pp. 698–699 (abstract).
Sicardy B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.
Sinclair AT (1978) The orbits of the satellites of Mars. *Vistas in Astronomy* 22: 133–137.
Singer SF (1968) The origin of the Moon and geophysical consequences. *Geophysical Journal of the Royal Astronomical Society* 15: 205–226.
Singer SF (1970) Origin of the Moon by capture and its consequences. *EOS, Transactions American Geophysical Union* 51: 637–641.
Singer SF (1972) Origin of the Moon by tidal capture and some geophysical consequences. *Moon* 5: 206–209.
Singer SF and Bandermann LW (1970) Where was the Moon formed? *Science* 170: 438–439.
Slattery WL, Benz W, and Cameron AGW (1992) Giant impacts on a primitive Uranus. *Icarus* 99: 167–174.
Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
Smith DE, Lemoine FG, and Zuber MT (1995) Simultaneous estimations of the masses of Mars, Phobos, and Deimos from distant spacecraft encounters. *Geophysical Research Letters* 22: 2171–2174.
Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.
Sohl F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo observations on the interior structure and chemistry of the Galilean satellites. *Icarus* 157: 104–119.
Solomon SC (1977) The relationship between crustal tectonics and internal evolution in the Moon and Mercury. *Physics of the Earth and Planetary Interiors* 15: 135–145.
Solomon SC (1986) On the early thermal state of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 435–452. Houston, TX: Lunar and Planetary Institute.
Soter S and Harris AW (1977) The equilibrium figures of Phobos and other small bodies. *Icarus* 30: 192–199.
Spohn T and Schubert G (2003) Oceans in the icy Galilean satellites of Jupiter? *Icarus* 161: 456–467.
Stephens DC and Noll KS (2006) Detection of six trans-neptunian binaries with NICMOS: A high fraction of binaries in the cold classical disk. *Astronomical Journal* 131: 1142–1148.
Stern SA, Bottke WF, and Levison HF (2003) Regarding the putative eccentricity of Charon's orbit. *Astronomical Journal* 125: 902–905.
Stern SA and McKinnon WB (2000) Triton's surface age and impactor population revisited in light of Kuiper Belt fluxes: Evidence for small Kuiper Belt objects and recent geological activity. *Astronomical Journal* 119: 945–952.
Stern SA, Parker JW, Duncan MJ, Snowdall JC Jr., and Levison HF (1994) Dynamical and observational constraints on satellites in the inner Pluto-Charon system. *Icarus* 108: 234–242.
Stern SA, McKinnon WB, and Lunine JI (1997) On the origin of Pluto, Charon, and the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 605–663. Tucson, AZ: University of Arizona Press.
Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.
Stevenson DJ (1984) The Uranus-Neptune dichotomy: The roll of giant impacts. In: *Lunar and Planetary Science XVII*, pp. 1011–1012.
Stevenson DJ (2001a) Origin of the Jovian Satellites. Paper presented at the *Jupiter Conference*, Boulder, CO, June 25–30.
Stevenson DJ (2001b) Jupiter and its moons. *Science* 294: 71–72.
Stevenson DJ, Harris AW, and Lunine JI (1986) Origin of satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 39–88. Tucson, AZ: University of Arizona Press.
Stone JM, Gammie CF, Balbus SA, and Hawley JF (2000) Transport processes in protostellar disks. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 589–601. Tucson, AZ: University of Arizona Press.
Szebehely V and Evans RT (1980) On the capture of the Moon. *Celestial Mechanics* 21: 259–264.
Szeto AMK (1983) Orbital evolution and origin of the Martian satellites. *Icarus* 55: 133–168.
Takata T and Stevenson DJ (1996) Despin mechanism for protogiant planets and ionization state of protogiant planetary disks. *Icarus* 123: 404–421.
Takeda T and Ida S (2001) Angular momentum transfer in a protolunar disk. *Astrophysical Journal* 560: 514–533.
Tanaka H, Takeuchi T, and Ward WR (2002) Three-dimensional interactions between a planet and an isothermal disk. I. Corotation and Lindblad torques and planet migration. *Astrophysical Journal* 565: 1257–1274.
Tanaka H and Ward WR (2004) Three-dimensional interaction between a planet and an isothermal gaseous disk. II. Eccentricity waves and bending waves. *Astrophysical Journal* 602: 388–395.
Tanigawa T, Ohtsuki K, and Machida MN (2012) Distribution of accreting gas and angular momentum onto circumplanetary disks. *Astrophysical Journal* 747: 47–63.
Taylor SR and Jakes P (1974) The geochemical evolution of the moon. In: Proceedings of the 5th Lunar Science Conference, pp. 1287–1305. Houston, TX: Lunar Science Institute.
Taylor DJ, McKeegan KD, and Harrison TM (2009) Lu-Hf zircon evidence for rapid lunar differentiation. *Earth and Planetary Science Letters* 279: 157–164.
Testi L, Natta A, Shepherd DS, and Wilner DJ (2003) Large grains in the disk of CQ Tau. *Astronomy and Astrophysics* 403: 323–338.
Tholen DJ, Buie MW, and Grundy WM (2012) *Asteroids, Comets, Meteors*, Proceedings of the conference held 16–20 May 2012 in Niigata, Japan. LPI Contribution 1667, id.6327.
Thomas PC (1993) Gravity, tides, and topography on small satellites and asteroids – Application to surface features of the Martian satellites. *Icarus* 105: 326–344.
Thomas PC (2010) Sizes, shapes, and derived properties of the Saturnian satellites after the Cassini nominal mission. *Icarus* 208: 395–401.
Thompson C and Stevenson DJ (1988) Gravitational instability in two-phase disks and the origin of the moon. *Astrophysical Journal* 333: 452–481.
Tolson RH, Duxbury TC, Born GH, et al. (1978) Viking first encounter of Phobos – Preliminary results. *Science* 199: 61–64.
Touboul M, Klien T, Bourdon B, Palme H, and Weiler R (2007) Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* 450: 1206–1209.
Touboul M, Puchtel IS, and Walker RJ (2012) ^{182}W evidence for long-term preservation of early mantle differentiation products. *Science* 335: 1065–1069.
Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
Touma J and Wisdom J (1994) Evolution of the Earth-Moon system. *Astronomical Journal* 108: 1943–1961.
Touma J and Wisdom J (1998) Resonances in the early evolution of the Earth-Moon system. *Astronomical Journal* 115: 1653–1663.
Tsiganis K, Gomes R, Morbidelli A, and Levison H (2005) Origin of the orbital architecture of the giant planets of the solar system. *Nature* 435: 459–461.
Turner NJ, Lee MH, and Sano T (2014) Magnetic coupling in the disks around young gas giant planets. *Astrophysical Journal* 783: 14, arXiv:1306.2276v1 [astro-ph.EP].
Veras D and Armitage PJ (2004) Outward migration of extrasolar planets to large orbital radii. *Monthly Notices of the Royal Astronomical Society* 347: 613–624.
Vokrouhlicky D, Nesvodrny D, and Levison HF (2008) Irregular satellite by exchange reaction. *Astronomical Journal* 136: 1463–1476.
Wada K, Kokubo E, and Makino J (2006) High-resolution simulations of a Moon-forming impact and post-impact evolution. *Astrophysical Journal* 638: 1180–1186.
Walsh KJ, Richardson DC, and Michel P (2008) Rotational breakup as the origin of small binary asteroids. *Nature* 454: 188–191.
Ward WR (1979) Present obliquity oscillations of Mars – Fourth-order accuracy in orbital E and I. *Journal of Geophysical Research* 84: 237–241.
Ward WR (1981) Orbital inclination of Iapetus and the rotation of the Laplacian plane. *Icarus* 46: 97–107.
Ward WR (1993) Density waves in the solar nebula – Planetesimal velocities. *Icarus* 106: 274–287.
Ward WR (1997) Protoplanet migration by nebula tides. *Icarus* 126: 261–281.
Ward WR (2002) *Astrophysical Tides: The Effects in the Solar and Exoplanetary Systems*. Nanjing, China: IAU Colloquium.

Ward WR (2009) Particle filtering by a planetary gap. In: *40th Lunar and Planetary Conference*, 1477.
Ward WR (2012) On the vertical structure of the protolunar disk. *Astrophysical Journal* 744: 11.
Ward WR and Cameron AGW (1978) Disc evolution within the Roche limit. *Lunar and Planetary Science Conference* 9: 1205–1207.
Ward WR and Canup RM (2000) Origin of the Moon's orbital inclination from resonant disk interactions. *Nature* 403: 741–743.
Ward WR and Canup RM (2013) The evection resonance and the angular momentum of the Earth-Moon system. *44th Lunar and Planetary Science Conference,* March 18–22, LPI Contribution 1719; p 3029.
Ward WR and Canup RM (2006) Forced resonant migration of Pluto's outer satellites by an impact-produced Charon. *Science* 313: 1107–1109.
Ward WR and Canup RM (2010) Circumplanetary disk formation. *Astronomical Journal* 140: 1168–1193.
Ward WR and Hahn JM (1994) Damping of orbital inclinations by bending waves. *Icarus* 110: 95–108.
Ward WR and Hahn JM (2000) Disk-planet interactions and the formation of planetary systems. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 1135–1158. Tucson, AZ: University of Arizona Press.
Weaver HA, Stern SA, Mutchler MJ, et al. (2006) Discovery of two new satellites of Pluto. *Nature* 439: 943–945.
Weidenschilling SJ (1977) Aerodynamics of solid bodies in the solar nebula. *Monthly Notices of the Royal Astronomical Society* 180: 57–70.
Weidenschilling SJ (1995) Can gravitation instability form planetesimals? *Icarus* 116: 433–435.
Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
Wetherill GW (1992) An alternative model for the formation of the asteroids. *Icarus* 100: 307–325.
Wetherill GW and Cox LP (1984) The range of validity of the two body approximation in models of terrestrial planet accumulation. I. Gravitational perturbations. *Icarus* 60: 40–55.
Wetherill GW and Cox LP (1985) The range of validity of the two body approximation in models of terrestrial planet accumulation. II. Gravitational cross sections and runaway accretion. *Icarus* 63: 290–303.
Wetherill GW and Stewart GR (1989) Accumulation of a swarm of small planetesimals. *Icarus* 77: 330–357.
Whipple FL (1964) The history of the solar system. *Proceedings of the National Academy of Sciences of the United States of America* 52: 565–594.
Widmer R, Masters G, and Gilbert F (1991) Spherically symmetric attenuation within the Earth from normal mode data. *Geophysical Journal International* 104: 541–553.
Wiechert U, Halliday AN, Lee D-C, Snyder GA, Taylor LA, and Rumble D (2001) Oxygen isotopes and the moon-forming giant impact. *Science* 294: 345–348.
Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
Willner K, Oberst J, Hussmann H, et al. (2010) Phobos control point network, rotation and shape. *Earth and Planetary Science Letters* 294: 541–546.
Wilner DJ, D'Alessio P, Calvet N, Claussen MJ, and Hartmann L (2005) Toward planetesimals in the disk around TW Hydrae: 3.5 Centimeter dust emission. *Astrophysical Journal* 626: L109–L112.
Wilson EH and Atreya SK (2004) Current state of modeling the photochemistry of Titan's mutually dependent atmosphere and ionosphere. *Journal of Geophysical Research* 109: E06002.
Winters RR and Malcuit RJ (1977) The lunar capture hypothesis revisited. *Moon* 17: 353–358.
Wisdom J (2004) Spin-orbit secondary resonance dynamics of Enceladus. *Astronomical Journal* 128: 484–491.
Wisdom J and Holman M (1991) Symplectic maps for the n-body problem. *Astronomical Journal* 102: 1528–1538.
Wood JA (1972) Fragments of terra rock in the Apollo 12 soil samples and a structural model of the moon. *Icarus* 16: 462–501.
Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites. *Science* 167: 602–604.
Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
Youdin AN and Goodman J (2005) Streaming instabilities in protoplanetary disks. *Astrophysical Journal* 620: 459–469.
Youdin A and Johansen A (2007) Protoplanetary disk turbulence driven by the streaming instability: Linear evolution and numerical methods. *Astrophysical Journal* 662: 613–626.
Youdin AM, Kratter KM, and Kenyon SJ (2012) Using Pluto's newest moon to constrain the masses of Nix and Hydra. *Astrophysical Journal* 755, Id 17, 11 pp.
Young EF and Binzel RP (1994) A new determination of radii and limb parameters for Pluto and Charon from mutual event lightcurves. *Icarus* 108: 219–224.
Yuan D-N, Sjogren WL, Konopliv AS, and Kucinskas AB (2001) Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research* 106: 23377–23401.
Yung YL, Allen M, and Pinto JP (1984) Photochemistry of the atmosphere of Titan – Comparison between model and observations. *Astrophysical Journal Supplement Series* 55: 465–506.
Zhang J, Dauphas N, Davis AM, Leya I, and Fedkin A (2012) The proto-Earth as a significant source of lunar material. *Nature Geoscience* 5: 251–255.
Zolotov MY, Owen T, Atreya S, Niemann HB, and Shock EL (2005) An endogenic origin of Titan's methane. *American Geophysical Union*, *Fall Meeting 2005*, abstract #P43B-04.

10.18 Interiors and Evolution of Icy Satellites

H Hussmann, DLR Institute of Planetary Research, Berlin, Germany
C Sotin, JPL, California Institute of Technology, Pasadena, CA, USA
JI Lunine, Cornell University, Ithaca, NY, USA

© 2015 Elsevier B.V. All rights reserved.

10.18.1	**Introduction**	605
10.18.2	**Spectroscopic Constraints on Composition**	606
10.18.3	**Elemental Abundance from Density**	608
10.18.4	**Size, Shape, and Mass**	609
10.18.5	**Modeling the Interior Structure**	610
10.18.5.1	Density and Composition	610
10.18.5.2	Gravity	610
10.18.5.3	Two- and Three-Layer Structural Models	611
10.18.5.4	Pressure-Induced Phase Transitions	613
10.18.5.5	Oceans	613
10.18.5.6	Thermal State	615
10.18.5.7	Heat Transfer	616
10.18.5.8	Effect of Volatiles	617
10.18.6	**Evolution of Satellite Interiors**	618
10.18.6.1	Tidal Heating	618
10.18.6.2	Heat Balance	619
10.18.6.3	Thermal Histories	619
10.18.7	**Interior Structure of Selected Icy Satellites**	620
10.18.7.1	Ganymede and Callisto	620
10.18.7.2	Europa	622
10.18.7.3	Titan	625
10.18.7.4	Enceladus	626
10.18.7.5	Rhea	628
10.18.7.6	Iapetus	628
10.18.7.7	Triton	629
10.18.8	**Future Prospects for Determining Satellite Internal Structure**	629
Acknowledgments		630
References		630

10.18.1 Introduction

The icy moons of the outer solar system are numerous and diverse in their properties. Although none of these objects approach the size of Earth, three – Ganymede, Callisto, and Titan – are larger in diameter than the planet Mercury. In this chapter, we focus on the large regular satellites thought to have coaccreted with their parent planets. An exception is Triton, which is a large satellite thought to have been captured by Neptune. The small irregular satellites (~60 known objects orbiting Jupiter and several tens of objects in orbit around the other giant planets) are not discussed here. Most of the latter are thought to have been captured by the giant planets.

Internal pressures for the large moons approach several GPa, sufficient to yield phase changes and significant compression in the water ice component – but not the rock. Water ice being among the major constituents of these bodies implies that the interior physics, in terms of the effect of the phase diagram and equation of state of water ice, is as complex as for the Earth, where analogously, megabar pressures compel rock to undergo compressive and phase change effects. The addition of soluble and generally more volatile compounds to the water ice lowers the melting temperature, yields partial melts, and greatly complicates the treatment of the interiors of icy satellites. As was first discussed by Lewis (1971), the large icy satellites may contain global liquid layers underneath their icy surfaces. The presence and the extent of such liquid layers will depend on the thermal state and composition of the H_2O layer in the outer part of the satellite. Since the solar insolation is only of minor importance for the internal heat budget of the satellites in the outer solar system, the energy required to form and maintain liquid layers has to be provided from the interior. Radiogenic heating in the rock component and – in some cases – tidal friction are the main heat sources for satellites. The presence of volatiles such as ammonia, significantly reducing the melting temperature of ice, will be crucial for the existence and maintenance of these internal oceans. Lewis (1971) used an equilibrium condition with the heat production due to radioactive nuclides being equal to the heat loss through the ice shell. However, the assumption of equilibrium may not be valid throughout the satellite's history, especially at the early stages of evolution. Time-dependent thermal histories and the

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00178-0

consequences for internal structure were calculated by Consolmagno and Lewis (1976, 1978) for the icy Galilean satellites and for several size classes of model satellites, respectively. These models, however, suffered from the lack of reliable data, especially for the mean densities deduced from the masses and the radii, of the outer planet satellites.

With the *Voyager* flybys (1979 Jupiter, 1981 Saturn, 1986 Uranus, and 1989 Neptune), data on the physical properties and images of the surfaces of the satellites improved significantly. The new constraints led to refinements of the then-available models (Ellsworth and Schubert, 1983; Lupo, 1982; Schubert et al., 1981, 1986) and, due to the observed diversity of surface features of the satellites, to the development of more complex models for individual satellites. By the end of the 1970s, the significance of tidal heating for some of the outer planet satellites became apparent (Peale et al., 1979), leading to considerable interest in thermal and interior structure models of Io, Europa, Ganymede, and Enceladus, the satellites for which tidal heating was considered to be the cause of volcanic activity (Io) and intense resurfacing in the satellites' history.

The *Galileo* mission (Jupiter orbiter, 1995–2003) has significantly improved our knowledge on the internal structure of the Galilean satellites (for a post-*Galileo* review, see Schubert et al., 2004). Until recently, the Galilean satellites were the only satellites in the outer solar system for which higher gravitational moments were determined from spacecraft flybys. Potential theory and the theory of equilibrium figures in combination with the acquired data allow for a more detailed inference of the internal structure of satellites if their rotational state and the tidal potentials are known (Dermott, 1979; Zharkov et al., 1985). *Galileo* data have been used to construct models of the four Galilean satellites (Anderson et al., 1996, 1998, 2001a,b; Schubert et al., 2004; Sohl et al., 2002), consisting of ice, rock, and metals. The data provide evidence for a considerable degree of internal differentiation for the three inner Galilean satellites. Callisto, however, is only partially differentiated (Anderson et al., 2001b), which may be a consequence of different accretion scenarios (Canup and Ward, 2002; Estrada and Mosqueira, 2006) and orbital evolution of Ganymede into the Laplace resonance, an orbital commensurability between the three inner Galilean satellites, Io, Europa, and Ganymede (Peale and Lee, 2002; Showman and Malhotra, 1997). Barr and Canup (2010) suggested different impact rates during late heavy bombardment (about 700 Ma after planet formation) as an alternative scenario to explain the differences between Ganymede's and Callisto's interior structure.

The acquisition of similar data has been under way since 2004 from the *Cassini*/*Huygens* mission for some of the icy satellites of Saturn. At the time of this writing, gravity data have been obtained from flybys at Titan, Enceladus, and Rhea. Among the results is an evidence for a liquid water ocean beneath Titan's crust (Iess et al., 2012) and partially differentiated structures for both Rhea and Titan (Iess et al., 2007, 2010), while the data for Enceladus do not as yet constrain the internal structure beyond the bulk density (Ducci et al., 2012; Rappaport et al., 2007). A handful of further gravity flybys are planned for the remainder of the solstice extended mission set to end in 2017. In addition, important information on the spin state of Titan has been laboriously obtained through *Cassini* RADAR data (Meriggiola and Iess, 2012) and will continue to be obtained up to the end of the mission.

In this chapter, we review the composition of icy satellites, describe the construction of interior structure models, present recent models composed of rock (including iron) and ice, describe how other materials affect the interior, and close with recent discoveries and key issues. We focus generally on the larger moons (**Figures 1** and **2**), those for which compression is important but porosity is not (**Figure 3**), and will briefly discuss some smaller moons, for example, Enceladus, where recent *Cassini* data have constrained particular properties of the interior.

10.18.2 Spectroscopic Constraints on Composition

With the exception of Enceladus, an active satellite, space probes do not have access to subsurface material of the icy moons. However, close inspection of surface materials by

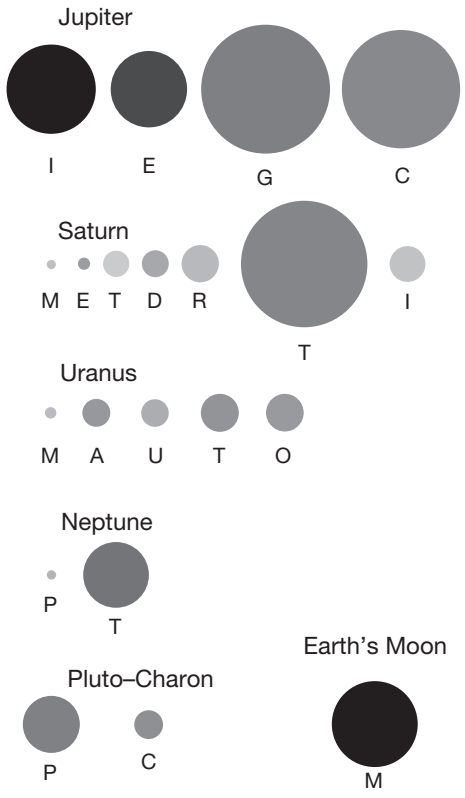


Figure 1 The largest moons of the solar system. Shown are the satellite systems of Jupiter, Saturn, Uranus, Neptune, and the Pluto–Charon system, including all moons with a radius larger than about 200 km. The smallest included here is Mimas with a radius of 198.8 km. Within each system, the satellites are ordered by the distance from the planet (distances are not to scale). The shading illustrates the mean density relative to each other, with Io being the densest (~3500 kg m^{-3}) and Tethys the one with the smallest density (956 kg m^{-3}). Pluto and the Moon, which is slightly smaller than Io, are included for comparison. From left to right and top to bottom: Io, Europa, Ganymede, Callisto; Mimas, Enceladus, Tethys, Dione, Rhea, Titan, Iapetus; Miranda, Ariel, Umbriel, Titania, Oberon; Proteus, Triton; Pluto, Charon; the Moon.

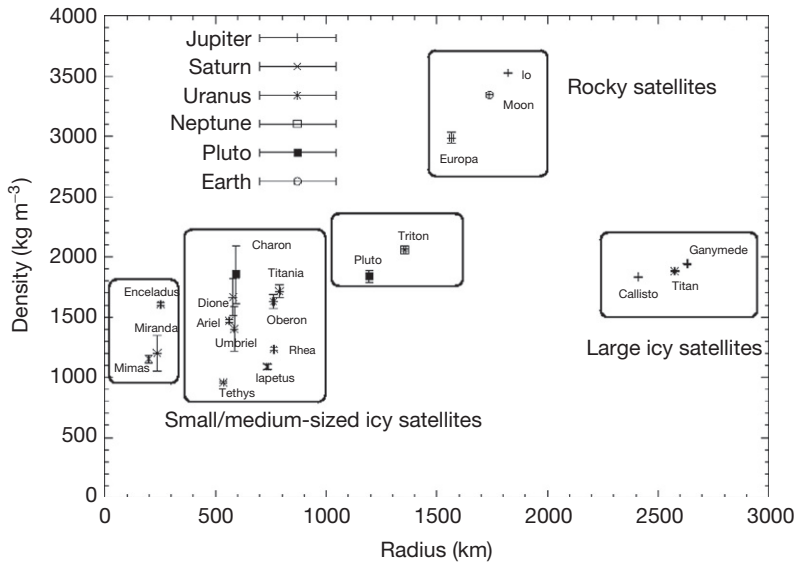


Figure 2 Mean density as a function of radius for all satellites larger than about 200 km and Pluto. Different symbols represent the respective primary planets. Several groups can be distinguished with respect to size and rock content. The small, medium-sized, and large icy satellites all contain substantial amounts of ices. Pluto and Triton are distinct from the medium-sized icy satellites with respect to size and their relatively high rock content. Europa, the Moon, and Io are dominated by rock. Modified from Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.

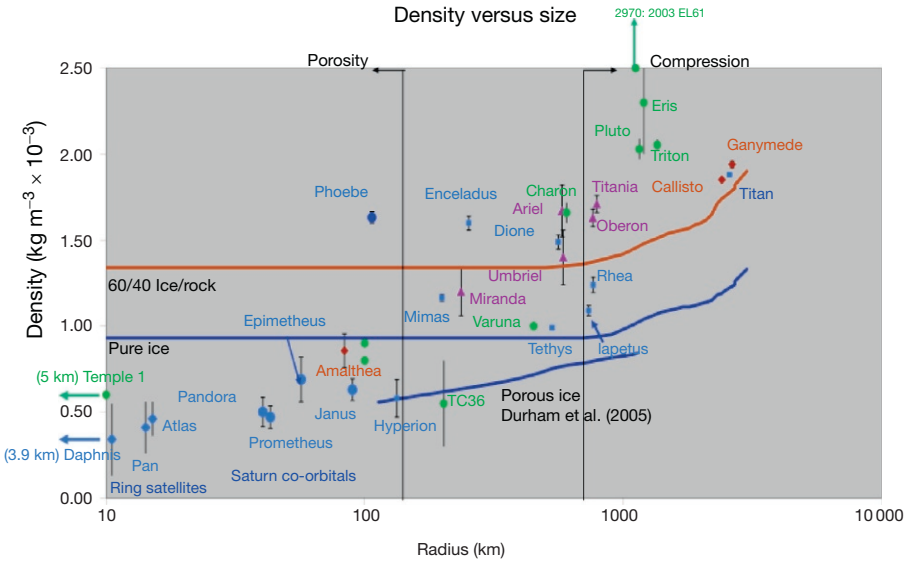


Figure 3 Bulk (measured) density of icy satellites and selected Kuiper belt objects versus radius. Shown also as vertical lines are rough limits on the size above which compressional effects alter the bulk density from the material density of the objects and below which porosity becomes potentially important in lowering the density. The two approximately horizontal lines show run of model density versus radius for water ice and silicate bodies containing 60% water ice (top line) and 100% water ice (bottom line). Reproduced from Johnson TV and Estrada PR (2009) Origin of the Saturn system. In: Dougherty M, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini-Huygens*, pp. 55–74. Dordrecht: Springer.

remote-sensing instruments can reveal the nature of physical and chemical processes involving the surface, subsurface, and the satellite's environment. In this section, we briefly review the constraints on the composition of icy satellite surfaces derived from spectroscopy.

Water ice is the major carrier of oxygen in the solid material of the outer solar system and is abundant in the moons of the giant planets. Spectroscopic observations, mostly in the near-infrared, constrain the presence and amount of water ice at the surface and therefore also in the crust. Callisto's surface contains 20–45% by weight water ice (Calvin and Clark, 1991), while Ganymede's surface ice content is even larger at upward of 90%, and Europa's approaches 100% (Clark, 1980). Other oxygen-bearing compounds are present or suspected as well, such as hydrogen peroxide on Europa (Hand and Brown, 2013) and CO_2 on Ganymede (Hibbits et al., 2003). In the

Saturn system, the intermediate-sized satellites Mimas, Enceladus, Tethys, Dione, Rhea, and Hyperion show water ice absorption features on their surface (Filacchione et al, 2013), as does Phoebe (Clark et al., 2005), as the major – but not sole (Brown et al., 2006) – spectroscopically active component of their surfaces. *Cassini* measurements of plume material coming from the subsurface of Enceladus show strong evidence for subsurface liquid water (Postberg et al., 2011), a variety of non-water ice species including methane and CO_2 (Waite et al., 2006), and it is thus unclear whether the methane was present at the time of Enceladus' formation or formed from CO_2 (Mousis et al., 2009b). Titan's surface, beneath its haze-shrouded atmosphere, also contains water ice (Tomasko et al., 2005) but mostly covered with organic material (Clark et al., 2010). In the Uranian system, water ice is present on the five major satellites, and carbon dioxide has been detected on Ariel, Umbriel, and Titania (Cartwright et al., 2013). Water ice has been detected on Neptune's moon Triton along with nitrogen, methane, carbon dioxide, and carbon monoxide (Brown et al., 1995); Pluto; and its moon Charon (see **Chapter 10.20**).

10.18.3 Elemental Abundance from Density

The major solid-forming materials in the outer solar system, based on simple solar elemental abundance considerations of the relevant elements, are water ice, silicates, and iron compounds. These are the materials that we will call, for brevity's sake, respectively, 'ice' and 'rock.' Where we consider 'ices' other than water ice, for example, solid carbon dioxide or ammonia–water mixtures, we will specifically call these out. Likewise, we will consider the rock phase to explicitly include the corresponding solar abundances of iron and nickel, the primary metals. Only where required (specifically in the case of Ganymede, with its metal core distinct from the silicates), we will consider these separately. To first order, then, our models of icy satellites consist of ice and rock, including the iron component. Chondritic abundances are assumed for the composition of the rock and iron component.

The significance of the densities of the icy satellites lies in the constraints they impose on interior models, and in turn (for satellites that are subjected to strong compressional effects), these models are the intermediary between the measured densities and the corresponding rock-to-ice ratios within the satellites. Whether water dominates over rock, or vice versa, is a strong function of whether carbon is mostly locked up in oxidized or reduced molecules, because the solar elemental C/O ratio is close to one-half according to the most recent determinations (**Table 1**). The revised solar abundances reviewed in Asplund et al. (2009) render the ice-to-rock ratio fairly sensitive to the molecular state of the carbon compared to previous tabulations. However, a number of processes can alter the rock-to-ice ratio in outer planet satellites from the value predicted by solar abundance and a particular assumed value of oxidized carbon (which would be mostly carbon monoxide, but including carbon dioxide as well). The process of satellite disk formation may alter molecular abundances and selectively remove either ice or rock (Canup and Ward, 2006; Estrada and Mosqueira, 2006; see also **Chapter 10.17**). If the abundance of organics of high carbon number (and thus mostly refractory) is large enough to remove carbon from the system, then the relationship between water abundance and oxidized carbon species weakens considerably, as **Figure 4** shows.

This last case is particularly interesting because a variety of high-carbon-number organics were likely present in condensed phases in the solar system. Moderate reductions in the oxygen abundance in the inner part of the solar nebula, for example, associated with diffusive redistribution of water vapor and cold trapping of water ice (Stevenson and Lunine, 1988) will result in the gas-phase thermochemical production of acetylene, hydrogen cyanide, and other light organics (Cyr et al., 1999). Chemistry in nebular shocks, photochemistry on the surface of disks, or particle chemistry associated with cosmic rays that penetrate the disk will produce a variety of organic molecules with higher carbon-to-hydrogen ratio than methane. It is possible that we are seeing evidence of this chemistry in the detection of higher hydrocarbons in comets and organic phases in Phoebe, likely an object captured by Saturn from solar orbit (Clark et al., 2005; Johnson and Lunine, 2005). The large amount of carbon dioxide seen in each does not

Table 1 Solar photosphere abundances of relevant elements

O	0	±0.05
C	−0.26	±0.05
N	−0.86	±0.05
Si	−1.18	±0.03
Mg	−1.09	±0.04
Fe	−1.19	±0.04

Log_{10} (abundance) relative to oxygen = 0.
Source: Asplund M, Grevesse N, Sauval AJ, and Scott P (2009) The chemical composition of the Sun. *Annual Review of Astronomy and Astrophysics* 47: 481–522.

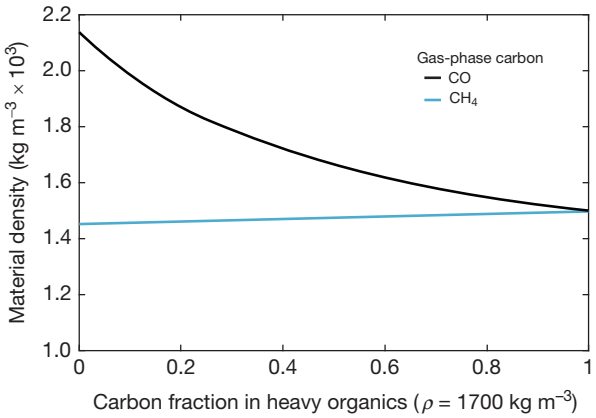


Figure 4 Plot of density of an uncompressed (small) rock–ice moon versus amount of carbon present in refractory organics. The refractory organics are assumed to have a mean density of 1700 kg m^{-3}, equivalent to amorphous carbon. The black line illustrates the case in which the rest of the carbon is in the form of carbon monoxide, while the shaded line illustrates the case in which it is in the form of methane. For a heavy organics fraction in excess of 0.5, the difference in density between the two cases is $<10\%$. Reproduced from Wong M, Lunine JI, Atreya SK, et al. (2008) Oxygen and other volatiles in the giant planets and their satellites. *Reviews in Mineralogy and Geochemistry* 68: 219–246.

necessarily mean that the latter dominated in abundance over carbon monoxide but may reflect instead the vast difference in volatility between the two. The effect of the presence of more refractory (carbon-rich) organics is to lower the sensitivity of the bulk density to the ratio of carbon monoxide (which consumes water) to methane.

In sum, while it is clear from spectroscopic data, chemical considerations, and satellite bulk densities that water ice is the dominant solid-forming material in the outer solar system, the relative abundances of other species – particularly carbon-bearing ones – remain uncertain and may be highly variable from object to object and from one satellite system to another (Dodson-Robinson et al., 2009). Densities and hence rock-to-ice ratios do not provide a direct one-to-one correspondence with particular carbon compositions and hence formation conditions. Further progress will require detailed spectroscopy of surfaces and the ability to measure in detail – in extraordinary cases such as Enceladus – the chemistry of material emitted from the interiors.

10.18.4 Size, Shape, and Mass

Typical radii of the major satellites range between about 200 km (Mimas and Miranda) and a few thousand kilometers with Ganymede, Titan, and Callisto being the largest satellites in the solar system (see **Figure 1** and **Table 2** for comparison of the sizes and masses).

With the exception of Titan, satellite surfaces can be observed directly by spacecraft imaging systems and ground-based telescopes, yielding the physical size of these objects that is required to determine the density. The radius of Titan's solid surface, buried beneath hundreds of kilometers of opaque haze, was first revealed by a radio occultation performed by *Voyager* 1 (Lindal et al., 1983). For irregular objects, which include moons as large as Jupiter's Amalthea and Saturn's Phoebe and Hyperion (the last with a long dimension exceeding 200 km), the determination of the total volume is more difficult and typically requires spacecraft observations at reasonable spatial resolution over a significant fraction of the body (e.g., Saturn's moon Phoebe; Porco et al, 2005).

More subtle is determining the shapes of moons, which reflect the satellites' internal structure under certain conditions (Dermott, 1979; Hubbard and Anderson, 1978; Zharkov et al., 1985). Moons tend to relax to a triaxial ellipsoidal shape determined by hydrostatic equilibrium of the rotating and tidally deformed body. This refers to the long-term fluid relaxation of a tidally distorted body. It must be distinguished from the tidal deformations due to periodic forces varying on timescales of the orbital period due to the orbital eccentricity or inclination, nonnegligible obliquity, or of a nonsynchronous rotational state. Deviations from hydrostatic equilibrium may occur, for instance, due to internal activity, incomplete relaxation, or a 'frozen in' structure that has been adjusted to equilibrium conditions different from the present ones. Measurements by *Cassini* in the Saturn system reveal at least one case (Iapetus) where the shape of the body is best represented by an oblate spheroid (Thomas, 2010; Thomas et al., 2007) inconsistent with the satellite's present-day rotation. It has been suggested (Castillo-Rogez et al., 2007) that the observed bulge, which did not adjust to the present equilibrium conditions, is due to a faster rotation rate in the satellite's past. In general, the shape of a satellite provides important clues to the orbital history of the body and to the rigidity over time of its internal layers.

Density and radial mass distribution inferred from the moment of inertia factor (see also **Chapter 10.02**) are the key parameters that constrain interior structure models. The mass of a satellite may be determined from two fundamentally different types of observations (Jacobson, 2004): visual (astrometric) observations of satellite motions from Earth-based telescopes or imaging systems on spacecraft and tracking via radio signal of the path of a spacecraft as it flies close by a moon and experiences the effect of the gravitational pull on its

Table 2 Characteristics of the largest icy satellites and Io

		ρ (kg m^{-3})	Radius (km)	Mass (10^{22} kg)	Eccentricity	P_{rot} (days)	Silicate mass (10^{22} kg)	%mass of ice	Power (TW)	MoI factor
J	Io	3525	1822	8.932	0.004	1.77	8.932	0	0.447	0.37824
	Europa	2989	1565	4.800	0.010	3.55	4.459	7.1	0.230	0.346
	Ganymede	1942	2631	14.819	0.0015	7.15	10.037	32.3	0.502	0.3115
	Callisto	1835	2410	10.760	0.007	16.70	6.835	36.5	0.342	0.3549
S	Titan	1881	2575	13.453	0.029	15.95	8.799	34.6	0.440	0.34
	Rhea	1237	763.5	0.231	0.001	4.52	0.062	73	0.003	0.372/0.391[a]
U	Titania	1715	788.9	0.3527	0.002	8.71	0.205	42	0.01	?
N	Triton	2061	1353.4	2.140	0	r 5.88	1.538	28	0.08	?
P	Charon	1759	603.6	0.162	<0.001	6.39	0.0976	40	0.005	?

The first column gives the planet (J for Jupiter, S for Saturn, U for Uranus, N for Neptune, and P for Pluto). To assess the radiogenic power at present time, we assume that the density of silicates is 3525 kg m^{-3}, a value equal to Io's density, an ice density of 1000 kg m^{-3}, and a radiogenic heating rate in the silicates of 5×10^{-12} W kg^{-1}, respectively. P_{rot} is the rotational period that is equal to the period of revolution for all the major satellites (synchronous state). In the case of Triton, 'r' indicates the retrograde orbit. Values for the Galilean satellites are taken from Schubert et al. (2004) and for Rhea from Iess et al. (2007). For Rhea, however, Schubert and Anderson (2007) derive an MoI value of 0.391. Values for Titan are taken from Iess et al. (2010). For the other satellites, see Hussmann et al. (2010). Mean densities, silicate mass fraction, and radiogenic heat production rate are calculated from the given values.

[a]Based on different analyses of the flyby data, there remains a discrepancy in the derived values for the moment of inertia of Rhea (see Anderson and Schubert, 2007, 2010; Iess et al., 2007).

trajectory. The latter is straightforward in terms of the nature of the effect: the stronger the satellite mass for a given flyby distance, the greater the effect on the spacecraft trajectory (measured in terms of the velocity perturbation). The astrometric observations depend on identifying mutual gravitational interactions among satellites or between satellites and the parent planet. The simplest effect – observing the wobble of a planet around the center of mass of a binary planet–moon system – provided the mass ratio of Pluto's moon Charon relative to Pluto (Olkin et al., 2003) and the system mass (Tholen and Buie, 1990). For giant planet systems, one looks for orbital commensurabilities that lead to orbital changes sensitive to satellite mass. For example, in the Saturn system, Mimas–Tethys and Enceladus–Dione resonances provide masses of these satellites, while small satellites at Lagrange points of larger moons also provide mass information on the latter (see Jacobson, 2004 for more details). Not all satellite masses can be well determined in this way, and tracking of spacecraft during close flybys remains the definitive mass-determination technique. Values for the Saturnian satellite system, for which four spacecraft have been used to determine densities, are given in Table 4 of Jacobson et al. (2006) and **Table 2**.

10.18.5 Modeling the Interior Structure

Due to the lack of measurements obtained by a seismic network, by deep surface penetration, or by ground-penetrating radar, interior structure models of the icy satellites are based on indirect evidence. The following characteristics are the main clues to investigate the structure of these moons: (1) radius and mass, (2) gravity field, (3) rotational state and shape, (4) magnetic field, (5) surface temperatures and heat flow, (6) composition of surface and atmosphere, and (7) activity at the surface. In addition to these features characterizing mainly a satellite's present state, knowledge on its evolution helps constrain possible internal states. This includes (1) surface geology and tectonics, which are telltale signs of present or past energetic processes; (2) orbital dynamics, being important for satellites evolving due to tides or in resonances; and (3) the chemical environment during accretion, that is, the composition and physical state of the solar nebula and planetary subnebula. The relevance of the items listed earlier varies for each individual satellite. However, in the following, we will illustrate how the combination of certain aspects can be used to gain a broad picture of the interiors of the icy moons.

10.18.5.1 Density and Composition

The satellite's radius R and its density ρ (or equivalently its radius and mass M if the shape is known or if spherical symmetry is assumed) are the basis of the models. To determine the distribution of mass within the icy satellites requires close spacecraft flybys, sometimes repeatedly, in order to resolve more precisely a body's gravitational field. The spacecraft will move along a trajectory that is perturbed by the various components of the body's gravitational field expressed as a spherical harmonic series (Hubbard and Anderson, 1978; Rappaport et al., 2001; Zharkov et al., 1985).

Assuming that the icy moons consist of rock (including iron) and nonporous ice, an assumption that is to first order reasonable on the basis of the compositional abundances of the outer solar nebula, the rock and ice mass fractions can be inferred from the mean density. Whereas the ice density is well constrained ($\sim$920 kg m^{-3}), the rock density may vary between about 2500 kg m^{-3} for hydrated rock (Schubert et al., 1986) and 8000 kg m^{-3} for pure iron. A typical value of nonhydrated silicate rock is 3500 kg m^{-3}, which will be used in the following as a reference value whenever we refer to rock. This value is close to the mean density of Io that is composed of rock and iron only, lacking a major component of volatiles. It is therefore used as a typical value for the rock/iron component of the other satellites.

For most of the satellites, it is unclear if they are differentiated into several distinct layers or homogeneous. Even for the same material, the density is pressure- and thus depth-dependent. In contrast to the larger terrestrial planets, however, compression within the rocky layers can be neglected and the rock densities can be regarded as constant. As shown by Sohl et al. (2002), even for the largest satellite, Ganymede, the density variations within each layer are small if self-compression is taken into account. However, the internal pressures can be sufficient to cause phase transitions from ice-I to high-pressure ice phases ice-II, ice-III, ice-V, and ice-VI, the densities of which (up to 1300 kg m^{-3}) significantly exceed the ice-I density (e.g., Schubert et al., 1986; Sotin et al., 1998).

10.18.5.2 Gravity

Besides the possibility to exactly determine the mass and radius of a satellite, close spacecraft flybys can provide measurements of the gravitational moments J_2 and C_{22}. The coefficient J_2 describes the polar oblateness mainly caused by the satellite's rotation and can best be determined with a polar flyby; C_{22} describes the equatorial bulge that is – in case of synchronous rotation – pointing toward the primary due to tidal interaction. It can best be determined with an equatorial flyby. According to first-order potential theory, these coefficients completely determine the satellite's moment of inertia C about the rotational axis, provided that the satellite is in hydrostatic equilibrium. C is defined as

$$C = \frac{8\pi}{3}\int_0^R \rho(r) r^4 \mathrm{d}r \qquad [1]$$

where $\rho(r)$ is the run of density with radial distance from the interior. **Table 2** gives basic parameters of selected icy moons, including C/MR^2. The determination of the dimensionless moment of inertia factor MoI $= C/MR^2$ yields an additional equation to infer the internal structure of a satellite that can be solved simultaneously with the mass balance equation. The lower the MoI factor is, the greater is the concentration of mass toward the center. For a homogeneous sphere, MoI $= 2/5$. In case of large satellites and planets, the MoI factor is slightly smaller than 2/5 even for homogeneous composition due to compression and phase changes at depth. The MoI factors of the Galilean satellites (see **Table 2**) were determined from flybys of the *Galileo* spacecraft and imply considerable degrees

of differentiation for Io (Anderson et al., 2001a), Europa (Anderson et al., 1998), and Ganymede (Anderson et al., 1996). Callisto, however, is regarded as partly differentiated (Anderson et al., 2001b).

Measurements of Titan's gravity field have been obtained by the *Cassini* spacecraft from various flybys. The mass (or more precisely the *GM* value), higher moments of the gravity field, the moment of inertia about the rotational axis, and the tidal potential Love number k_2 have been derived from Doppler tracking data (Iess et al., 2010, 2012). In a similar way, *Cassini* flybys were used to determine the static gravity fields of Rhea (Iess et al., 2007) and Enceladus (Ducci et al., 2012). In the case of Enceladus, data analysis from the flybys is still ongoing (as of August 2013).

Spacecraft flybys can be used to infer whether hydrostatic equilibrium is attained. The hydrostatic first-order theory yields $J_2 = 10/3\ C_{22}$, which can be verified if J_2 and C_{22} were independently determined from at least two close flybys. A ratio of 10/3 would be a strong argument in favor of hydrostatic equilibrium, although a combination of nonhydrostatic contributions resulting in the same ratio cannot be ruled out entirely.

In case of hydrostatic equilibrium, C_{22} is related to the moment of inertia through the secular (fluid) Love number k_f (e.g., Schubert et al., 2004) depending on the internal density distribution. k_f describes the complete relaxation of the body on long timescales (i.e., the fluid response) adjusting to the given tidal and rotational potentials. For a homogeneous body, $k_f = 3/2$, independent from its density. The relation of k_f, generally depending on the density distribution within the satellite, to the gravitational coefficient C_{22} can thus be used to constrain the satellite's internal structure. For a synchronously rotating satellite, to first order (e.g., Schubert et al., 2004),

$$C_{22} = \frac{1}{4} k_f \frac{\omega^2 R^3}{GM} \qquad [2]$$

where ω is the angular frequency of the orbital and rotational period and G is the gravitational constant. For the satellite's axial moment of inertia C, it follows (e.g., Kaula, 1968)

$$\frac{C}{MR^2} = \frac{2}{3}\left[1 - \frac{2}{5}\left(\frac{4 - k_f}{1 + k_f}\right)^{1/2}\right] \qquad [3]$$

In hydrostatic equilibrium, the extent of the satellite's internal differentiation – whether the ice has separated from the rock or the rock from the metal, thereby segregating into layers according to the density – can be thus obtained from C_{22}. Using eqn [1] and the value of C, the density distribution within a satellite can be constrained. However, due to the inherent ambiguity of the gravity data, such solutions are not unique (see **Chapters 10.02** and **10.05**).

Albeit not proving it, the ratio of C_{22} and J_2 may provide strong evidence if a moon actually is in hydrostatic equilibrium. Such measurements were successfully performed by the *Galileo* spacecraft at Io, the only Galilean satellite where the tidal and rotational responses could clearly be separated. Satisfying the $J_2 = 10/3\ C_{22}$ constraint, Io is very likely in hydrostatic equilibrium (Anderson et al., 2001a). The results are less conclusive for the other Galilean satellites, in particular for Callisto. As a consequence of the equatorial flyby, geometries J_2 and C_{22} could not be determined independently (Schubert et al., 2004). In the cases of Titan and Rhea, the quadrupole moments are consistent with a hydrostatic state (Iess et al., 2007, 2010). Although nonhydrostatic contributions cannot be ruled out completely (even when $J_2 = 10/3\ C_{22}$), the hydrostatic theory is frequently used as a first approach to construct interior structure models of the icy satellites, especially for the large ones.

The moments of inertia about the principal axes can also be derived from the rotational state of a satellite. One way suitable for rapidly rotating bodies is to determine the precession of the spin pole. The precession rate depends on all three moments of inertia and can be used to further constrain the principal moment about the rotation axis in combination with the quadrupole moments of the gravity field. In this case, the assumption of hydrostatic equilibrium is not needed. However, the precise determination of the precession rate is difficult because of the long timescales of precession. In case of synchronously rotating satellites, the determination of the satellite's obliquity can also be used to determine the moments of inertia. In case the satellite obtains a so-called *Cassini state*, the spin pole adjusts its angular separation from the orbit pole so that it remains coplanar with the orbit pole and the invariant pole, as the former precesses about the latter (Bills and Nimmo, 2008). If both the gravity parameters J_2 and C_{22} and the obliquity are known, then the moment of inertia can be determined without the assumption of hydrostatic equilibrium. However, to apply this theory, it has to be shown that the satellite actually has obtained one of the equilibrium states, known as *Cassini states*.

10.18.5.3 Two- and Three-Layer Structural Models

As an illustration, the two-layer model, consisting of a rock core and a H_2O layer, will be discussed (see also **Chapter 10.02**). Such a model may be a reasonable approach for the medium-sized satellites of Saturn and Uranus, in case of full differentiation. Larger satellites will undergo phase transitions from ice-I to high-pressure ice phases. It should also be noted that even for the medium-sized satellites, for example, for Rhea, the phase transition from ice-I to ice-II can occur in the central region, if the satellite is not fully differentiated in ice and rock (Anderson and Schubert, 2007; Castillo-Rogez, 2006; Schubert et al., 1986). However, in case of full differentiation, the pressures in the outer ice layer are sufficiently small for the ice-I phase to remain stable.

From a simple mass balance equation for a spherical body, the radius R_c and the mass M_c of the core can be directly obtained if the densities are known:

$$R_c = R\left(\frac{\rho - \rho_m}{\rho_c - \rho_m}\right)^{1/3}, \quad M_c = 4/3\pi R_c^3 \rho_c \qquad [4]$$

where ρ_c and ρ_m are the densities of the core and mantle (in this case ice-I), respectively. Results for possible core sizes are shown in **Figure 5** for the medium-sized satellites of Saturn and Uranus. Strikingly, there is no decrease of rock content with increasing distance from the planet in the Saturnian system. This stands in contrast to the Jovian system, in which the observed density gradient is consistent with a radial

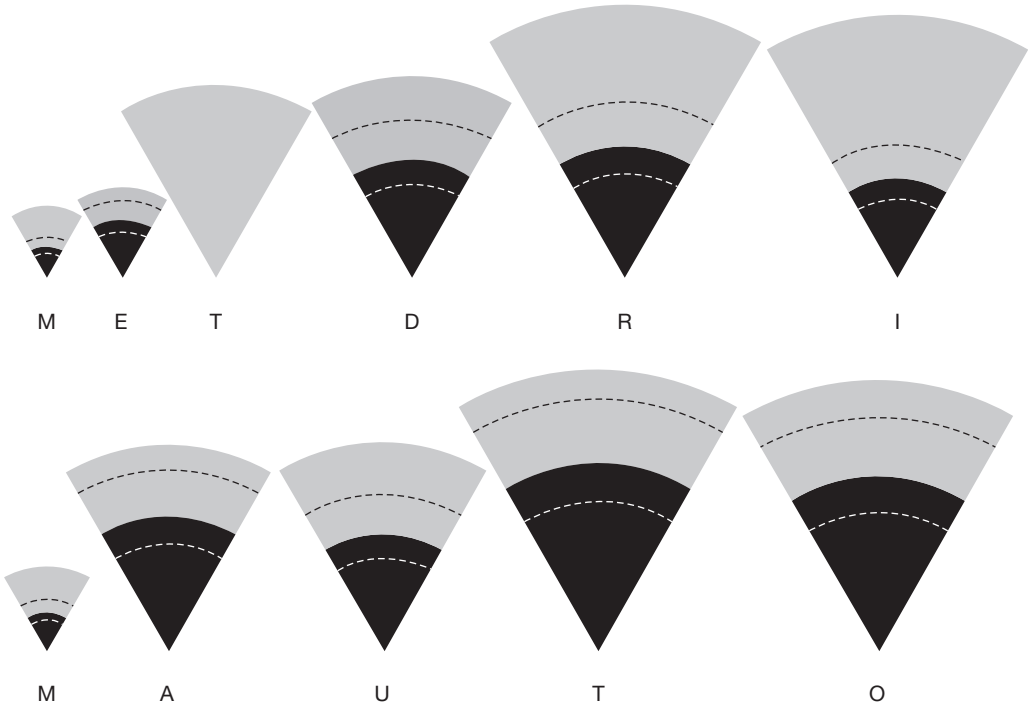


Figure 5 Possible internal structures of the midsized icy satellites of Saturn and Uranus consistent with the mean density are shown assuming complete differentiation into a rock core ($\rho_c = 3500$ kg m^{-3}) and an icy mantle ($\rho_m = 960$ kg m^{-3}). Complete differentiation is assumed here, only to illustrate the rock content. Whether differentiation was actually completed or if the satellites consist of homogeneous ice–rock mixtures has to be inferred from spacecraft flybys. For Rhea, for instance, complete differentiation can be ruled out on the basis of *Cassini* data (Anderson and Schubert, 2007; Iess et al., 2007). From left to right and bottom to top: Mimas, Enceladus, Tethys, Dione, Rhea, Iapetus; Miranda, Ariel, Umbriel, Titania, and Oberon. Sizes are to scale. The dashed lines show the possible range of core sizes for a variation of the rock density between 2000 and 6000 kg m^{-3}, corresponding to hydrated rock (larger cores) and iron-rich rock (smaller cores), respectively. The strong variation of rock content in the Saturnian system is remarkable. Tethys is essentially an icy sphere, whereas the rock-mass fraction and thus core sizes differ considerably for the other satellites. Additionally, there is no obvious correlation of rock content with increasing distance from Saturn. It should be noted that prior to the determination of the higher moments of the gravitational fields, homogeneous ice–rock mixtures cannot be ruled out. The latter would imply ice-II phases in the central regions of the largest of the midsized satellites, Titania, Oberon, and Rhea.

temperature distribution in the Jovian subnebula. It is also noteworthy that with a mean density of only 956 kg m^{-3}, Tethys is lacking a substantial rock-mass fraction, whereas the moons in its vicinity, Enceladus and Dione, contain the highest rock-mass fractions (55% and 48%, respectively) of the midsized Saturnian satellites. The origin of such an irregular distribution that may imply collisional disruption in the inner part of the Saturn system (Mosqueira et al., 2010) is still an open question. In the Uranian system, variations in the rock content are smaller, ranging between 40% (Ariel) and 60% (Titania). An exception is Miranda (23%), which, however, is significantly smaller than the other satellites of Uranus and had a unique origin (Greenberg et al., 1991). The bulk rock-mass fraction is increasing as we go outward from Saturn to Uranus, a trend that continues if the Neptunian satellite Triton (74%) and the Pluto–Charon system (70% and 63%, respectively) are included.

For the two-layer model and more generally for the N-layer model, we may additionally use the moment of inertia (provided it is determined from flybys) to constrain the internal structure. Counting the layers from inside out, the mass balance equation and the equation for the moment of inertia have to be solved simultaneously:

$$\begin{aligned} M &= 4/3\pi\left[\rho_1 R_1^3 + \sum_{j=2}^{N}\rho_j\left(R_j^3 - R_{j-1}^3\right)\right] \\ \mathrm{MoI} &= 2/5\frac{1}{\rho R^5}\left(\rho_1 R_1^5 + \sum_{j=2}^{N}\rho_j\left(R_j^5 - R_{j-1}^5\right)\right) \end{aligned} \quad [5]$$

In this set, the density and the radius of each layer are in principle unknown, except for the outermost layer for which the radius coincides with the total radius ($R_N = R$). Thus, only two of the $2N-1$ unknown quantities can be obtained from eqn [5], implying an inherent ambiguity of multilayered interior structure models. Further assumptions, for example, on the layer densities are thus required to construct the models. In eqn [5], we assumed that the density within each layer is constant and thus neglected compressional effects – an assumption justified even for the largest icy satellites (Sohl et al., 2002).

The Jovian satellite Europa is a three-layer-type example, consisting of an iron core, a rocky mantle, and an outer H_2O layer (Anderson et al., 1998; Kuskov and Kronrod, 2005; Schubert et al., 2009; Sohl et al., 2002). For Europa, the moment of inertia was determined from *Galileo* flybys (see **Table 2**) using the assumption of hydrostatic equilibrium,

and thus, the set of equations mentioned earlier can be fully applied. Provided that the densities of the H_2O layer and the rocky layer are known, the set of eqn [5] can be solved, yielding the density and the radius of the core, respectively. Depending on its composition and thus density, the core can be as large as about 50% of Europa's total radius. The outer H_2O shell can have an extension of about 80–170 km (Anderson et al., 1998). A typical result for Europa's structure is shown in Figure 6. However, on the basis of the current data, an H_2O shell on top of a uniform mixture of rock and metals, although more unlikely, cannot be ruled out completely (Anderson et al., 1998; Schubert et al., 2009).

10.18.5.4 Pressure-Induced Phase Transitions

For a given density profile of a spherical satellite, it is straightforward to calculate the pressure $P(r)$ as a function of the radial distance r from the satellite's center solving the hydrostatic equation:

$$dP = -g(r)\rho(r)dr \qquad [6]$$

where g is the gravitational acceleration. Besides compression effects, which are of minor importance, the internal pressure is crucial for the possible occurrence of phase transitions within the ice or rock layers.

The internal structures of the largest icy satellites, Ganymede, Titan, Callisto, Europa, and Triton are complex because (1) there is evidence at least for the Galilean satellites that they contain iron cores, (2) pressures are sufficiently high for phase transitions of water ice to occur (at least for Ganymede, Callisto, and Titan), and (3) internal heat sources may be sufficient to maintain liquid water oceans underneath the ice-I shell (Spohn and Schubert, 2003).

Up to a pressure of about 0.2 GPa, ice-I will be stable, transforming into ice-III, ice-V, and ice-VI with increasing pressure. In Figure 7, the phase diagram for pure H_2O is shown. For a comprehensive discussion on high-pressure polymorphs of ices and their physical properties, see, for example, Hobbs (1974). A typical interior structure considering the phase transitions and corresponding densities of the high-pressure ice phases is shown as an example in application to Ganymede in Figure 8.

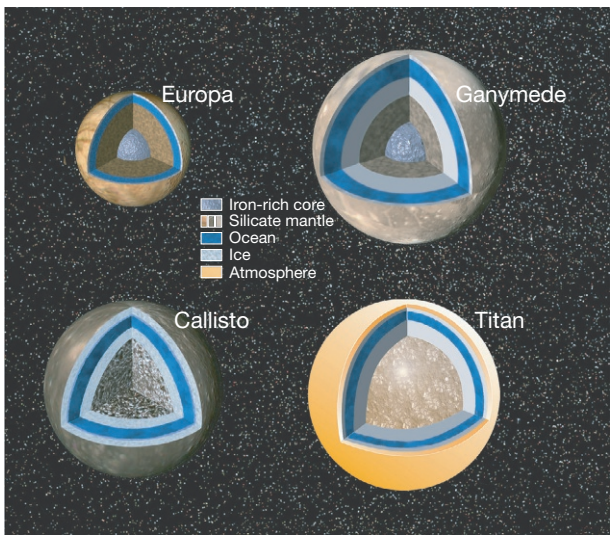


Figure 6 Models of the internal structures of large icy satellites and Europa (adapted from Sotin C and Tobie G (2004) Internal structure and dynamics of the large icy satellites. *Comptes Rendus De L'Academie Des Sciences Physique* 5: 769–780.)

10.18.5.5 Oceans

The fact that the ice-I melting temperature decreases with increasing pressure allows for the existence of liquid layers between the ice-I shell and the high-pressure ice phases, if the internal temperature is in the right range (see Figure 9). Volatiles, like ammonia (NH_3), are important because even small amounts substantially reduce the melting temperature of water ice (Figure 10).

Since the early work of Lewis (1971), the existence of subsurface water oceans in icy satellites and the relevance of ammonia (or other volatiles) were discussed, especially for the largest satellites. Because of possible tidal heating, the Jovian satellite Europa has been investigated in substantial detail. No concurrence with respect to the existence of an ocean was reached in the earlier models (Cassen et al., 1979, 1980; Ojakangas and Stevenson, 1989; Ross and Schubert, 1987; Squyres et al., 1983b) because of the unknown efficiency of convection in the ice shell. However, the existence of such an ocean is presently considered as the most likely case for Europa. It is consistent with thermal modeling (Hussmann et al., 2002; McKinnon, 1999; Mitri and Showman, 2005; Spohn and Schubert, 2003; Tobie et al., 2003, 2005) and with the geologic record (Carr et al., 1998; Pappalardo et al., 1998, 1999, and references therein), it explains the formation of cycloidal features on the surface (Hoppa et al., 1999), and it is consistent with the induced magnetic field detected at Europa (Kivelson et al., 2000).

Induced magnetic fields are generated, because the moons, which revolve around Jupiter in the planet's equatorial plane, are subject to the time-varying magnetic field of Jupiter. At the

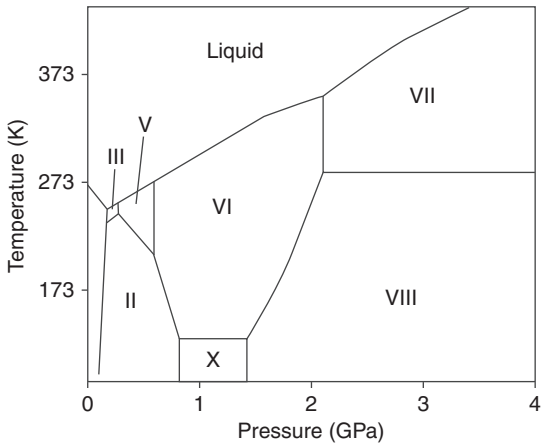


Figure 7 Phase diagram of pure water ice. The different high-pressure ice phases are shown. Most relevant for the icy satellites are ice-I, ice-II, ice-III, ice-V, and ice-VI.

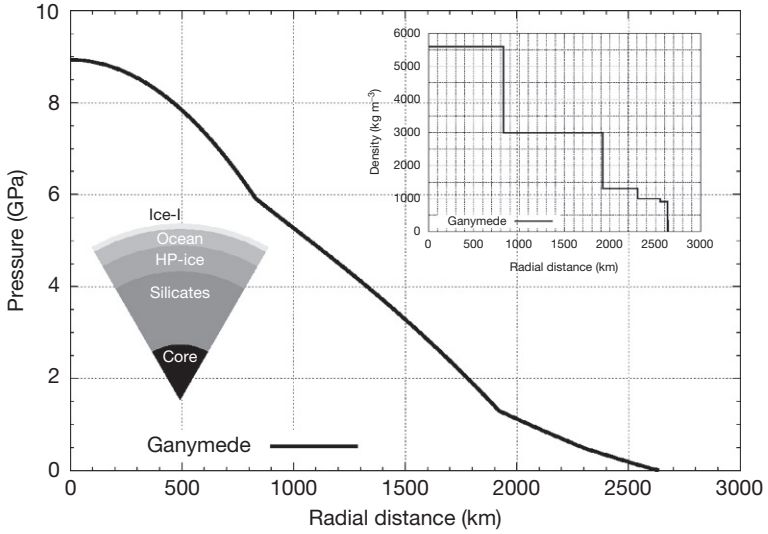


Figure 8 Interior structure model of Ganymede as a type example for a large icy satellite. Shown is the pressure and corresponding density (upper right) as a function of the radial distance from the planetary center. Based on the moment of inertia and mass balance constraints, the internal structure is calculated assuming hydrostatic and thermal equilibrium (equivalent to the Titan models described in Sohl et al., 2003). Further evidence for the differentiation of an iron core comes from the proof of Ganymede's intrinsic magnetic field during the *Galileo* mission.

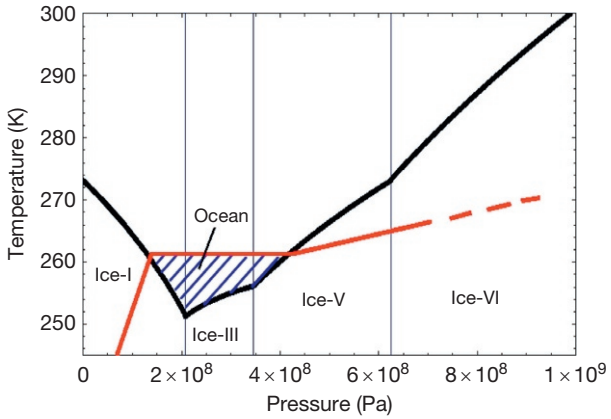


Figure 9 Conceptual sketch of oceans in icy moons. The melting curve of ice-I is shown in black (compare also with **Figure 7**) as a function of internal pressure (and thus with increasing depth). The red curve, representing a possible temperature profile, intersects the melting curve at two points. In between, the solid ice phase will not be stable and a liquid layer will form, in this case between the ice-I and the ice-V layers. The blue area shows the range where oceans are possible, even for lower temperatures. The higher the temperatures, the greater is the vertical extension of the ocean. The minimum temperature allowing for oceans is about 250 K in case of pure water ice. It may be substantially smaller with the presence of ammonia in the ocean (not shown here). The vertical lines mark the locations of the phase transitions.

location of the moons, the amplitude and the direction of the components of the Jovian field vary on a timescale of Jupiter's rotational period (9.925 h), because of an angle of about 9° between Jupiter's rotation axis and the axis of the Jovian magnetic field. Due to these variations, an electric current and consequently a secondary (or induced) magnetic field will be generated in electrically conducting regions within the moons. The amplitude and the direction of the induced field will be correlated with the variations of the Jovian principal field. Such fields were detected at Callisto, Europa, and possibly at Ganymede (Khurana et al., 1998; Kivelson et al., 2000, 2002; Neubauer, 1999; Zimmer et al., 2000). For Ganymede, the data analysis is more complicated because Ganymede additionally has an intrinsic magnetic field (Kivelson et al., 2002). The data analysis shows that the induced fields are generated close (a few hundred kilometers or less) to the surface, ruling out the possibility of currents in the deep interior (e.g., within metallic cores) as a cause for the induced fields (Zimmer et al., 2000). The most probable explanation for the induced fields being generated close to the surface is the electrolytic conductivity due to the presence of salts within a liquid water layer (Khurana et al., 1998; Neubauer, 1999). The induced magnetic fields are extremely important because it is so far the only observational evidence for the present-day existence of oceans. Surface features, for instance, may have an age of several My or more and may thus not necessarily represent the ocean's present state. The magnitude of the induction response is proportional to the product of electrical conductivity and thickness of the ocean. Estimates of the salinity of the ocean could therefore be used to constrain the ocean thickness (e.g., Hand and Chyba, 2007).

Liquid layers can also be expected in other large icy satellites (Lewis, 1971), especially at Titan (Grasset et al., 2000; Sohl et al., 1995, 2003), Ganymede, Callisto (Spohn and Schubert, 2003), and Triton (Hussmann et al., 2006; Lewis, 1971; Ruiz, 2003), provided that there are antifreezes, for example, ammonia, present reducing the melting temperature of water ice. The presence of ammonia can be expected on the basis of models of the solar nebula and condensation temperatures and theories of satellite formation at least for the Saturn system and beyond (Lewis, 1972). The abundance expected for the icy satellites, based on models for the Saturnian subnebula and on data obtained from the composition of comets and clouds around young stellar objects, is of the order of a few percent (Bockelée-Morvan, et al., 2004; Dartois and d'Hendecourt, 2001; Lunine and Stevenson, 1987; Mousis et al., 2009a).

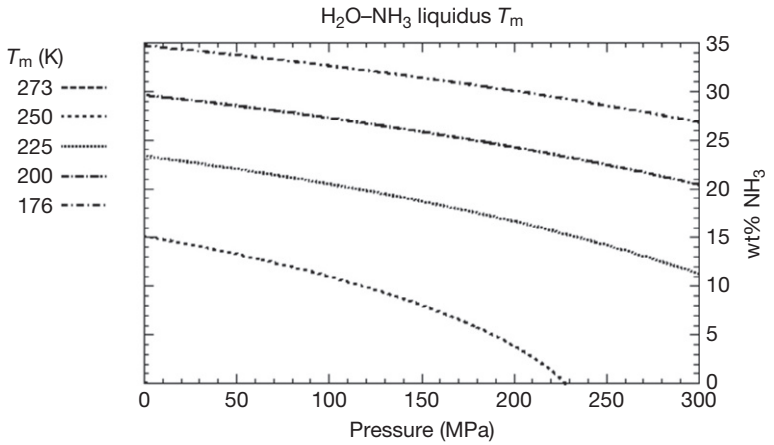


Figure 10 Contours of the melting temperature as a function of ammonia concentration and pressure for an H_2O–NH_3 mixture according to the parameterization of Leliwa-Kopystyński et al. (2002). Reproduced from Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.

For the largest satellites, Ganymede, Titan, and Callisto, internal oceans are expected to be located between the ice-I layer and the high-pressure ice layers. Differentiated Triton and Pluto models imply a pressure regime where the presence of high-pressure phases can be neither ruled out nor confirmed because of the inherent ambiguity of the models. A structure similar to the one of Europa, with a water–ammonia ocean being in direct contact with the silicate layer, is, however, conceivable for Triton and Pluto.

For the midsized satellites, present-day oceans could only exist in the largest members of this class of objects under certain conditions. Besides Ganymede, Titan, Callisto, and Europa, the best candidates for present-day subsurface oceans are Triton, Pluto, Titania, Oberon, and Rhea, provided that they are differentiated and contain ammonia (Hussmann et al., 2006). However, modeling based on gravity data from the *Cassini* mission suggests that Rhea is not fully differentiated (Anderson and Schubert, 2007; Iess et al., 2007) ruling out the possibility of an internal ocean. Rhea's structure is then likely to be an ice–rock mixture including ice-II in its central region (Castillo-Rogez, 2006). Although it is substantially smaller, Enceladus may also be added to the list of possible ocean candidates because of its strong endogenic and surface activity (Porco et al., 2006; see also details in **Section 10.18.7.4**).

The precise determination of the gravity field may provide an indirect way to detect subsurface oceans, due to the greater deformation of the outermost ice-I layer as a response to tidal forces, in case of an ocean. For a moon on an eccentric orbit, the gravitational force exerted by the parent planet varies in a predictable way and it is possible, with flybys at apoapsis and periapsis (farthest and closest distance to the parent planet) or in orbit, to obtain the time-varying tidal potential. The latter is a combination of the external tidal potential (caused by the planet) and the internal or secondary tidal potential caused by the redistribution of mass inside the satellite (i.e., due to the formation of a tidal bulge). The amplitude and phase lag of the satellite's tidal bulge as a response to the periodic external force can be used to determine the overall rigidity of the interior – which in turn constrains the possible presence of interior liquid layers. A liquid layer will decouple the outer ice shell from the deep interior, allowing for larger deformation rates of the ice shell due to the tidal forcing as compared to the case where an entirely solid ice shell is directly connected to the rock. Generally, tidal amplitudes will be about an order of magnitude larger in the presence of a liquid layer (Moore and Schubert, 2000, 2003; Sohl et al., 1995). The tidal potential and the tidal deformation of a satellite surface can tell whether a subsurface exists. Due to the linear dependence of the tidal response on the ice rigidity, a combination of independent measurements would be desired to constrain the ice thickness further. Promising strategies would be to combine measurements of the tidal potential, the tidal deformation, and the magnetic induction response in orbit around the satellite. In addition the determination of libration amplitudes (small periodic changes in the rotation rate due to tidal torques) can constrain the ice thickness in case of a de-coupling ocean layer (e.g., Van Hoolst et al. (2008) in application to Europa). The impact of the liquid layer on the tidal response of Titan's ice shell has provided indirect evidence for a subsurface ocean during the *Cassini* mission (Iess et al., 2012). If, and how, the tidal response of the ice shell can be used to infer the thickness of the ice shell and therefore the depth in which liquid water may be present has been investigated for possible future Europa missions (Wahr et al., 2006).

10.18.5.6 Thermal State

The interior structure is linked to a satellite's thermal state via the heat production rate in the rock component due to long-lived radiogenic isotopes (see also **Chapter 10.08**) and – in some cases – tidal heating. Assuming an elementary abundance of uranium, thorium, and potassium, consistent with the chondritic composition, the heating rate can be obtained from the laboratory-derived half-lives and the energy release per decay of the individual isotopes. The specific present-day heating rate H in the rocky component of the icy satellites is of the order of several 10^{-12} W kg^{-1}, which is roughly one order of magnitude less than the heating rate at the very beginning of the solar system. Here, and in the following, we neglect short-lived

isotopes that may have played a role at the early stages of solar system evolution, especially for planetesimals and their remnants, the asteroids and other small objects (Merk et al., 2002). The role of short-lived isotopes (mainly ^{26}Al, ^{60}Fe, and ^{53}Mn) has been recognized in various applications for the early evolution of satellites as a means of initiating phases of geologic activity (e.g., Barr and Canup, 2008; Castillo-Rogez et al., 2009).

From a satellite's rock-mass fraction m_{rock}, its overall present-day heat production is given by $H_{\text{total}} = Hm_{\text{rock}}$ (see Table 6 in Hussmann et al., 2010 for radiogenic heat production rates of solar system satellites). Since the heat transfer rate Q through the ice shell will depend on its thickness and physical state, the latter can be related to the heat production rate, determined by the chosen density and size of the rock core, if a steady-state $H_{\text{total}} = Q$ is assumed. Although such an equilibrium state may not be exactly matched, it is a reasonable approach for the icy satellites at late stages of their evolution, after primordial heat sources due to accretion, core formation, or early tidal heating have become negligible. However, it should be noted that for the Earth – the only planetary body for which reliable data have been obtained – only 50% of the surface heat flux can be attributed to radiogenic heating. The corresponding values are unknown for other planets and satellites, and therefore, the equilibrium condition must be regarded as a first assumption. Because of their small size and therefore greater surface to volume ratio (in comparison with the Earth and other terrestrial planets), the icy satellites cool faster and approximation to equilibrium is more likely for their present states. Deviations due to additional contributing heat sources, and possible disequilibrium between internal heating and the output of heat through the surface, however, cannot be ruled out completely.

The conductive heat flow through an ice shell of thickness D and temperatures T_0 and T_s at the base and the surface of the ice shell, respectively, are given by

$$Q = 4\pi R^2 k \frac{T_0 - T_s}{D}(1 - D/R) \qquad [7]$$

assuming a constant value for the thermal conductivity, generally being temperature-dependent. Considering a wide temperature range of 40–273 K, k varies for ice-I between 14 and 2 $\text{Wm}^{-1}\,\text{K}^{-1}$, according to the parameterization of Klinger (1981), $k(T) = 567/T$, which is valid for temperatures above 30 K. Similar values are obtained using the parameterization of Hobbs (1974). The factor $(1 - D/R)$ takes into account the spherical geometry and may be neglected for large satellites. The corresponding temperature profile is given by

$$T(r) = T_s + \frac{T_0 - T_s}{1 - R_0/R}\left(\frac{R_0}{r} - \frac{R_0}{R}\right) \qquad [8]$$

with $R_0 = R - D$ the radius of the base of the ice shell. Both equations are connected through Fourier's law. Here, the thermal conductivity and the surface temperature are taken to be constant with the latter ranging from about 100 K for the Jupiter system to about 40 K for Triton and the Pluto–Charon system. The values for individual satellites depend on surface characteristics, for example, albedo, or the presence of an atmosphere (e.g., Titan's surface temperature = 94 K, significantly exceeding the value (about 80 K) expected without an atmosphere). Setting the heat production into equilibrium with the heat flow, $Q = H_{\text{total}}$, the temperature at the base of the ice shell is given by

$$T_0 = T_s + \frac{DH_{\text{total}}}{4\pi R^2 k(1 - D/R)} \qquad [9]$$

The equations mentioned earlier are valid for purely conductive shells, only. The heat transfer through a thick and relatively warm ice shell, however, may be dominated by the much more efficient process of subsolidus convection. Whereas, at present, heat is mainly transferred by conduction in small-sized and midsized icy satellites, thermal convection can be the dominant process in large icy satellites.

10.18.5.7 Heat Transfer

Thermal convection is discussed in detail in **Chapter 10.08**. Whereas the basic equations describing convection in an ice shell are the same as for mantle convection, the physical properties, especially the temperature range in which ices are subject to viscous flow, are completely different as compared to rock. The viscosity, which is difficult to describe by first principles or by parameterizations based on laboratory measurements, is a key parameter controlling convection. It is strongly temperature-dependent, and different parameters have to be applied for various creep mechanisms operating in specific temperature and stress regimes. Furthermore, the viscosity is pressure- and grain size-dependent. The physical properties for H_2O ice, creep mechanisms, and laboratory data with regard to the dynamics of icy satellites are discussed, for example, in Weertman (1983), Durham and Stern (2001), Goldsby and Kohlstedt (2001), Sotin et al. (1998), McKinnon (2006), and Durham et al. (2010).

Due to the large viscosity contrast between the base of the ice shell and the surface, the icy satellites are thought to operate in the stagnant-lid regime (Solomatov, 1995; see also **Chapter 10.08**), characterized by a convective region and a relatively thick lid, in which the viscosity is too large for flows to occur. In the lid, the heat transfer is purely conductive and can be calculated by the equations given earlier. In cases of high heat flux, icy satellites may also operate in a 'mobile-lid' regime where convective plumes can rise close to the surface. These exceptional cases, characterized by a high heat flux, may be relevant for present Enceladus (Barr, 2008; Han et al., 2012). A schematic temperature profile for a satellite in the stagnant-lid regime is shown in **Figure 11**. The convective region is characterized by a small temperature drop, corresponding to a viscosity contrast of about an order of magnitude only. Since pressure changes are small within the ice shell, adiabatic temperature changes can be neglected, and except for the boundary layers, the temperature T in the well-mixed convective region can be regarded as constant. Several scaling laws, based on laboratory data and numerical experiments, have been proposed to describe both the heat flow and the onset of convection (Davaille and Jaupart, 1993; Deschamps and Sotin, 2001; Grasset and Parmentier, 1998; Solomatov, 1995). Non-Newtonian viscosities and the dependence on grain sizes for strongly temperature-dependent convection

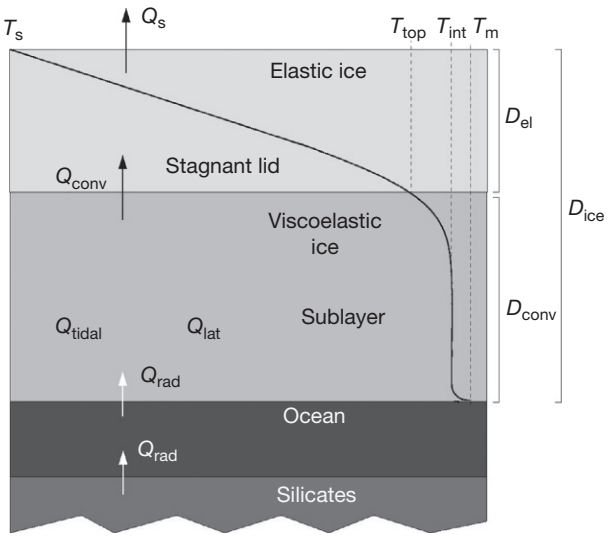


Figure 11 Schematic sketch of a convective ice-I layer at the surface of an icy satellite. Q_{rad} is the radiogenic heating rate from the silicate mantle. Q_{conv} is the heat flow out of the convective region, and Q_s the surface heat flow. Q_{lat} and Q_{tidal} may be additional heat sources in the ice if latent heat is released due to crystallization or in cases where tidal dissipation is an important factor. The ice-I layer is divided in the convective sublayer and the stagnant lid. The temperature (black curve) is high and, apart from the boundary layers at the very top and bottom, almost constant within the convective region. In the stagnant lid, the temperature is decreasing strongly corresponding to the conductive heat transport. The temperature profile can be constructed if the melting temperature T_m, that is, the temperature at the base of the ice shell and the surface temperature T_s, as well as material parameters, for example, the thermal conductivity, are known. Such a structure (not to scale) with a rocky mantle and an H_2O layer on top of it would be appropriate for the satellite Europa. Modified from Hussmann H, Spohn T, and Wiezcerkowski K (2002) Thermal equilibrium states of Europa's ice shell: Implications for internal ocean thickness and surface heat flow. *Icarus* 156: 143–151.

have been discussed, for example, by Barr et al. (2004), McKinnon (2006), and Barr and McKinnon (2007).

The thickness of the stagnant lid (and implicitly the thickness of the convecting sublayer) depends very sensitively on the value of the derivative of viscosity as a function of temperature, evaluated at the interior temperature in the sublayer, and on the value of the local Rayleigh number evaluated at the sublayer temperature (Solomatov and Moresi, 2000). This implies that in contrast to previous models (e.g., Schubert et al., 1979), the upper bound of the convective region is not determined by a fixed isotherm, but is found to be dependent on the temperature within the convective region itself (Davaille and Jaupart, 1993; Grasset and Parmentier, 1998). Using such an approach, the temperature at the top of the convective layer can be calculated as a function of the internal temperature determining also the corresponding heat flow out of the convective region if a viscosity law and an appropriate Rayleigh number are defined. In a steady state, the heat flow out of the convective region will be equal to the purely conductive heat flow through the stagnant lid, which can be used to determine the thicknesses of the individual layers. The more vigorous the convection, the thicker will be the convective region and the thinner the stagnant lid (for fixed temperatures at the surface and the base of the ice shell). At present, convection is expected to play a major role in the ice shells of the largest satellites, for example, Ganymede, Titan, Callisto, and Europa.

10.18.5.8 Effect of Volatiles

Independent of the mode of heat transport (convection or conduction), it has to be verified whether a solid ice shell could be stable under the given circumstances, that is, whether T_0 is above or below the melting temperature of ice-I. The determination of the pressure P_0 at the base of the shell is straightforward since the mass of the latter is known for a specified thickness D with the remaining mass being included in the sphere below the ice shell. The melting temperature $T_m(P_0)$ can then be determined from the laboratory-derived parameterizations for the given pressure. The reduction of the melting temperature due to a specified small amount of ammonia may be taken into account at this stage (see **Figure 10**). Parameterizations are given by Chizhov (1993) for pure H_2O and by Hogenboom et al. (1997), Leliwa-Kopystyński et al. (2002), Grasset and Pargamin (2005), and Choukroun and Grasset (2010) for H_2O–NH_3 mixtures. If T_0 exceeds the melting temperature, the solid ice-I phase will not be stable and a liquid water layer (ocean) will form underneath the ice shell (**Figure 9**). The upper bound of the ocean will be determined by the melting curve and can be calculated from the corresponding pressure. If T_0 is smaller than the melting temperature, the ice-I layer will be stable. In this case, it has to be verified whether or not the pressure of the ice-I/ice-III or ice-I/ice-II phase transition is reached. If not, the ice shell will consist of ice-I only, and the remaining bulk composition of the satellite will be rock, including the iron-rich components. If the pressure is sufficiently high for the phase change to occur, the lower part of the H_2O layer will consist of high-pressure ice covering the rocky interior.

As mentioned earlier, a liquid layer may form between the ice-I shell and the high-pressure phase. Such an ocean is assumed to be isothermal because the adiabatic temperature increase with increasing pressure is small for the icy satellites. Some applications, however, require that the adiabatic increase in ocean temperature (and density) be included to properly calculate radial expansion of icy satellites (e.g., Bland et al., 2009; Showman et al., 1997 in application to Ganymede). In a simple approach, the temperature at the base of the ocean can be equated with the temperature at the base of the ice shell, and the corresponding pressure (and hence depth) at the base of the ocean can again be calculated from the melting curve, this time for the high-pressure phase. Assuming a mean density of the high-pressure ice phase and the rock core, the extension of the latter is determined from the mass balance constraints. By varying the core size, and thus the heat production in the satellite's deep interior, an equilibrium model can be derived by an iterative process.

Depending on the number of available constraints, there may be some ambiguity inherent in the model results, and equilibrium may be attained for various core sizes, densities, and ice thicknesses. In **Figure 8**, the interior structure of Ganymede is shown as a typical example of a class of models consistent with the mass balance and moment of

inertia constraints assuming thermal equilibrium. Further constraints, for example, the presence of a magnetic field, can be used to restrict possible values of core sizes and densities (e.g., assuming the density of the Fe–FeS eutectic for the core).

10.18.6 Evolution of Satellite Interiors

The presence of subsurface oceans could have been an important factor in the satellites' histories, even for the smaller satellites, because of the greater radiogenic heating rate in the past. Thermal evolution models can be calculated from heat balance equations applying the same fundamental principles as for the terrestrial planets (see **Chapter 10.08**). The major modification in application to the icy satellites is the replacement of the physical properties of mantle rock by those of ice. Further complexity arises in models in which subsurface oceans play a role. Additional factors, for example, the release of latent heat due to crystallization of an ice layer and the increase of ammonia concentration in the subsurface ocean due to freezing of pure ice, have to be taken into account.

10.18.6.1 Tidal Heating

Besides radiogenic heating, additional heat sources are relevant especially at the beginning of satellite evolution. The most important ones are accretional heating, release of potential energy due to differentiation, and tidal heating (e.g., Hussmann et al., 2010; Schubert et al., 1986). The latter has been very efficient during the phase in which the satellites were not yet in a synchronous rotation state. This state was obtained later in the evolution by all large and midsized icy satellites. For synchronously rotating satellites, tidal heating will be most effective for (1) large satellites on (2) eccentric orbits (3) close to the primary with (4) high internal temperatures (see eqn [10] in the succeeding text). These conditions are best fulfilled by the Jovian satellite Io, for which tidal friction still is the dominating heat source (see **Chapter 10.08**). At present, tidal heating significantly contributes to the heat budget of Europa and Enceladus. The example of Enceladus shows that the internal composition and the thermal state of a satellite could be crucial for tidal effects. With respect to size, distance from the planet, and orbital eccentricity, we would expect tidal heating of Mimas instead of Enceladus. That Enceladus, and not Mimas, is internally active may be due to the higher rock content – and thus larger radiogenic heating – or due to different thermal–orbital histories of the two satellites (see **Section 10.18.7.4**).

Tidal heating may additionally have played a significant role in the early stages of evolution for several other moons involving highly eccentric orbits or nonsynchronous rotation. Ganymede and Triton, both showing a great diversity of surface terrain and tectonic features, are prominent examples for a specific orbital history, most likely involving strong tidal heating (McKinnon and Leith, 1995; Showman and Malhotra, 1997).

Even for synchronously rotating satellites, tidal friction plays a crucial role for the histories of some moons because it provides a link between the orbital evolution (because tidal friction in the satellite implies a loss in orbital energy and reduces the eccentricity) and the thermal state of a satellite (internal heating because of friction within the satellite). For a satellite in synchronous rotation with negligible inclination and obliquity, the tidal heating rate is given by (Segatz et al., 1988)

$$H_{\text{tidal}} = -\frac{21}{2}\frac{R^5 n^5 e^2}{G}\text{Im}(k_2) \qquad [10]$$

where n and e are the mean motion and orbital eccentricity of the satellite, respectively. $\text{Im}(k_2)$ is the imaginary part of the second-degree potential Love number that depends on the structure (density profile), tidally effective rheology (e.g., temperature-dependent viscosity of the planetary material), and frequency of the forcing (i.e., the mean motion n). Here, k_2 describes the viscoelastic response of the satellite to the external periodic forcing on the timescale of the orbital period – it has to be distinguished from the previously introduced fluid Love number, which describes a completely relaxed state of a planetary body on long timescales.

The complex body tide Love numbers are obtained by numerical integration of the linearized field equations for small viscoelastic deformations in a spherical and self-gravitating body subject to specific boundary and continuity conditions (e.g., Segatz et al., 1988). The correspondence principle that relates the viscoelastic problem to the well-known elastic problem permits to solve the resultant set of differential equations for the complex radial functions k_2(r) (potential), h_2(r) (radial displacement), and l_2(r) (tangential displacement) in the frequency domain (Zschau, 1978). In application to the deformation problem of satellites under periodic tidal forcing, layers with constant density and constant material parameters are usually assumed. The rheological model describes the relation between stress and strain taking into account the deformation of ice and rock as a function of frequency. The simplest assumption is the Maxwell model in which the deformation is elastic (i.e., energy-conserving) on timescales shorter than the Maxwell time, the ratio of tidally effective viscosity and rigidity. On timescales on the order of the Maxwell time, the deformation is viscoelastic and highly dissipative. On timescales significantly exceeding the Maxwell time, the response can be regarded as fluid. The Maxwell model is completely determined by two material parameters, viscosity and rigidity, that are generally temperature-dependent. More realistic rheological models (e.g., the Burgers model) include also transient parameters that are, however, difficult to constrain for the tidal problem by laboratory measurements. Equation [10] is independent from the rheological model chosen. This approach links the dissipation rate to the rheological model in a consistent way, and it is thus more general than the often used approximation of $\text{Im}(k_2) \sim k_2/Q$, where Q is a constant quality factor describing the ratio of the maximum elastic energy stored in one tidal cycle over the energy dissipated in that cycle. Tidal dissipation can be extremely relevant for the evolution of satellites because it connects the thermal state with the orbital and/or rotational state.

The link between tidal heating and thermal evolution has been intensely investigated in application to the Galilean satellites because of Io's thermal activity and the resonance locking of the inner three satellites (Fischer and Spohn, 1990; Greenberg, 1982; Ojakangas and Stevenson, 1986; Peale, 1986; Peale et al., 1979; Spohn, 1997; Squyres et al., 1983a; Yoder, 1979; Yoder and Peale, 1981; see also **Chapter 10.08**).

With respect to the icy satellites, tidal heating has been considered mainly in application to Europa to infer the thermal state and thickness of the ice shell (Cassen et al., 1979, 1980; Hussmann et al., 2002; Mitri and Showman, 2005; Ojakangas and Stevenson, 1989; Ross and Schubert, 1987; Sotin et al., 2009; Tobie et al., 2003; see also the details in **Section 10.18.7.2**) and in application to Enceladus to explain the tectonic surface features (Peale, 2003; Poirier et al., 1983; Ross and Schubert, 1989; Squyres et al., 1983a).

Thermal–orbital models have also been applied to explain Ganymede's tectonic surface features by an evolution into the Laplace resonance (Showman and Malhotra, 1997) and to investigate the thermal–orbital history of Europa and its ocean over long timescales (Hussmann and Spohn, 2004). The relation between the orbital history and tidal dissipation in Titan has been discussed by Sohl et al. (1995) and Tobie et al. (2005, 2006).

10.18.6.2 Heat Balance

The evolutionary paths of the icy moons are as diverse as is the appearance of their surfaces and their bulk structure. It is therefore impossible to describe all aspects by a simple set of governing equations. However, the basis of all evolution models is an energy balance between the internal heat sources and the heat transport equations depending on viscosity (in case of convection) and material parameters. The distribution of heat sources will depend on the state of differentiation and, in case of tidal heating, on the orbital state of the satellite. The initial conditions will furthermore depend on accretional processes and the thermal environment due to solar insolation. The heat balance equation for a spherical shell with density ρ, heat capacity c, temperature T, and volume V, bounded by the surfaces S_1 and S_2 ($S_1 > S_2$), is given by (Schubert et al., 1986)

$$\int_V \left(\frac{\mathrm{d}}{\mathrm{d}t}(\rho c T) - H_V \right) \mathrm{d}V = \int_{S_2} q_2 \mathrm{d}S - \int_{S_1} q_1 \mathrm{d}S \qquad [11]$$

where H_V is the total heat production rate per unit volume and q_1 and q_2 are the heat fluxes (power per unit surface) into the shell (through surface S_1) and out of the shell (through surface S_2), respectively. Appropriate equations describing the convective or conductive heat flows have to be applied to calculate q_1 and q_2. Both parameterized convection models and numerical approaches solving directly the Navier–Stokes equations for the convective ice shell subject to certain boundary conditions have been applied to investigate the evolution of icy satellites. Two-dimensional numerical models were extensively developed in application to the ice shells of the Galilean satellites (Barr and Showman, 2009; Showman and Han, 2004). The disadvantage of such an approach, in which fewer parameters are required, is an immense increase in computational time (further more for three-dimensional models) and the greater effort that is required to couple such models with other theories, for example, tidal heating and orbital evolution. Additionally, in some specific problems where the information on flow patterns as a result of numerical models is not needed, parameterized models may be preferred.

An integration of eqn [11] can be done for several shells with different material parameters or heat sources, describing the individual layers of the satellite. The initial temperature profile required for the integration is usually given by an assumed accretional profile of the satellite (e.g., Schubert et al., 1986).

10.18.6.3 Thermal Histories

Thermal histories have been calculated mainly in application to the Galilean satellites and the satellites of Saturn (Consolmagno and Lewis, 1976, 1978; Ellsworth and Schubert, 1983; Lupo, 1982; Multhaup and Spohn, 2007; Schubert et al., 1981, 1986). Typical evolution scenarios of homogeneous midsized icy satellites can be divided into four phases (Ellsworth and Schubert, 1983; Multhaup and Spohn, 2007; Schubert et al., 1986): (1) the initial accretional state is characterized by a cold interior, a temperature increase with radial distance from the center, and a steep thermal gradient at the very surface due, in part, to impact heating (e.g., Schubert et al., 1986); (2) the satellite is then heated from within due to radioactive decay of long-lived isotopes. The internal temperature increases and, after a typical time of the order of several tens of My, the temperature profile is reversed, with the central region being warmer than the outer part of the satellite; (3) depending on the heat sources available and the size of the satellite, convection may start while the satellite heats up. In this phase, maximum temperatures will be reached, with the satellite being in its most dynamic phase, characterized by a maximum surface heat flow and internal activity. Convection leads to more efficient cooling of the satellite whose temperature then again decreases with time for several Gy; (4) at a certain time, convection will stop due to the generally decreasing radiogenic heating rate, and the satellite will further cool conductively up to 4.5 Gy, that is, the present time.

According to such a scenario, all the small and midsized satellites are expected to be in the fourth, relatively cold phase, cooling conductively. However, further modifications of that scenario can arise if the internal temperatures exceed the melting temperature of ice-I or ice-I/volatile mixtures within the satellite. In that case, liquid layers may form and the satellite will differentiate, forming a rock core and an outer ice layer, possibly contaminated with rocks. A liquid layer may persist enclosed between the rock and the ice layers. Whether these layers can survive up to the present time will depend on the rock content (and thus heating rate of the satellite) and on the amount of volatiles included in the liquid. Because of freezing of pure ice at the top of the liquid layer, the volatile content in the liquid will increase, thereby impeding crystallization until the eutectic composition is reached. Further cooling will lead to the crystallization of the eutectic mixture. Lewis (1971) noted the significance of ammonia as the most important antifreeze component for icy satellites and concluded that the largest objects still may contain liquid layers in their interiors, a result that is consistent with more recent models of icy satellites (Hussmann et al., 2006; Spohn and Schubert, 2003).

It should be noted that in these evolution models, additional heat sources, like tidal heating, that may have been important for individual satellites are neglected. The enormous internal heating and activity of Enceladus (Porco et al., 2006;

see also the details in Section 10.18.7.4) shows that individual satellite histories may be completely different from the general scenario described here. However, most of the midsized satellites' surfaces are heavily cratered showing only little or no signs of recent activity, which is consistent with the general concept of evolution of this class of objects.

The importance of tidal heating for satellites was realized prior to the *Voyager* arrivals at Jupiter (Peale et al., 1979). Whereas radiogenic heating is the major heat source in the inner solar system, tidal heating has been recognized as an important driver for the evolution of satellites in the outer solar system in various applications. A notable difference to radiogenic heating is the strong feedbacks that can be initiated by tidal heating. Through the temperature-dependent rheological parameters, both the dissipation rate itself and the efficiency of heat transport can be affected by tidal heating. Furthermore, the energy that is dissipated comes from orbital and/or rotational energy of the satellite. Therefore, tidal heating connects the thermal evolution with the orbital and rotational histories of the satellites. In addition, it may have played a crucial role for geologic processes that have shaped the satellite surfaces. Thus, processes like tidal heating, convection, orbital evolution, geologic activity, and furthermore formation, differentiation, and later evolution of satellites are interrelated and interdependent. In case of orbital resonances, even the evolution of different satellites and their respective energy budgets are related to each other. We will discuss examples in applications to specific satellites further below.

For the large icy satellites, more complex evolution scenarios, including core formation or the segregation of rocks within the ice shells, have been applied to understand their physical characteristics. This includes models of the incomplete differentiation of Callisto to explain the substantial differences between Ganymede and Callisto (Barr and Canup, 2010; Kirk and Stevenson, 1987; McKinnon, 2006; Mueller and McKinnon, 1988; Nagel et al., 2004) or the differentiation and evolution of Titan (Fortes, 2012; Grasset and Sotin, 1996; Grasset et al., 2000). The three large icy satellites Ganymede, Callisto, and Titan are further discussed in the following section.

Besides the physical evolution described so far, the chemical history of the icy satellites, which is directly linked to the composition (including minor components) of the nebula in which the satellites accreted, plays an important role (e.g., Kuskov and Kronrod, 2005; Lunine and Stevenson, 1982). Models have been developed mainly in application to Titan to understand the unique coupling between interior, surface activity, and atmosphere of this icy moon (Lunine and Stevenson, 1987; Tobie et al., 2006; see also **Section 10.18.7.3**).

10.18.7 Interior Structure of Selected Icy Satellites

While the densities of the Galilean satellites and Titan were determined to high accuracy during the *Voyager* flybys, it was the *Galileo* spacecraft that obtained most of the information on the interior structure of the Galilean satellites as it orbited around Jupiter between 1995 and 2003. *Cassini* has provided important data during a set of dedicated 'radio science' flybys of Titan and other satellites in the Saturn system. In the following, we discuss a selection of the icy satellites in detail, focusing on results from the *Galileo* and *Cassini* missions with respect to internal structure. For the Saturn system, we discuss the largest satellites, Titan, Rhea, and Iapetus, as well as the presently active satellite Enceladus. Saturn's midsized satellites are also discussed by Matson et al. (2009). In addition to the satellites of Jupiter and Saturn, we discuss Neptune's moon Triton, for which activity at the surface has been revealed by the *Voyager* 2 spacecraft.

10.18.7.1 Ganymede and Callisto

Ganymede is the largest satellite in the solar system with a radius of 2634 km. Its density of 1940 kg m^{-3} lies in between the densities of silicates and that of water. The observation that Ganymede has its own intrinsic magnetic field and forms a small magnetosphere inside Jupiter's vast magnetosphere was a major discovery of the *Galileo* mission. Ganymede has a clear signature of a permanent dipole (Kivelson et al., 1998). As pointed out by Stevenson (2003), a permanent magnetism explanation is conceivable (Crary and Bagenal, 1998) but unlikely, and the most reasonable interpretation is a dynamo in the metallic core (Schubert et al., 1996). Kivelson et al. (2002) suggest a dipole moment equal to 1.32×10^{20} A m^2. The dipole is tilted by 176°, almost antiparallel to the spin axis. The few flybys allow only for the determination of the degree 1 and 2 coefficients of the spherical harmonic decomposition of the magnetic field. By determining the loss of power between degrees 1 and 2, the radius of a liquid iron core has been evaluated around 200 km (Kivelson et al., 2002). By comparison with Earth, a liquid Fe–S core is expected although the nature of the light element in Earth's core is still debated. Because of the self-sustained magnetic field, we can assume that within Ganymede, the metal-rich component has separated from the rock, forming an iron core, surrounded by a silicate mantle. The outer layers consist of H_2O, which will be present in the form of high-pressure ice, ice-I, or liquid water, depending on internal pressure, temperature, and composition as a function of depth.

A dipolar field cannot fully explain the magnetic data. Two models have been proposed by Kivelson et al. (2002) in order to explain the remaining signal: quadrupole terms resulting from the dynamics of the core dynamo and an induced magnetic dipole driven by the time-varying component of the externally imposed magnetic field of Jupiter's magnetosphere. Both models give equally satisfactory fits to the data although the inductive field is favored (Kivelson et al., 2002) due to its presence within Europa and Callisto (see the succeeding text).

The gravity data alone can provide the moment of inertia assuming hydrostatic equilibrium. A value of 0.3115 ± 0.0028 has been obtained (Anderson et al., 1996; Schubert et al., 2004), implying that Ganymede is strongly differentiated (**Figures 6** and **8**). A two-layer model (silicate core and H_2O mantle) cannot explain this very small moment of inertia if reasonable values are taken for the densities of silicate and H_2O (Anderson et al., 1996). It should be noted that there are nonhydrostatic contributions to the gravity field of Ganymede. Those contributions can be explained by mass anomalies that might correlate with the geologic units (dark vs. bright terrain) on Ganymede (Palguta et al., 2009).

An iron-rich core that is required to explain the small value of the moment of inertia is in complete agreement with the interpretation of the magnetic data (Bland et al., 2008; Kimura et al., 2009). However, the radius of each interface cannot be precisely determined since it depends on the choice made for the densities (see also **Chapter 10.02** on the inherent ambiguity of such models). The thickness of the H_2O layer is in the range between 600 and 900 km (Deschamps and Sotin, 2001; Sohl et al., 2002). Due to the high-pressure value reached at the bottom of the H_2O layer, high-pressure ice phases must exist below depth of roughly 150 km. The H_2O layer is subdivided into ice-Ih, ice-III, ice-V, and ice-VI sublayers (**Figure 8**).

The state of the H_2O outer layer is a subject of debate. If a liquid layer exists, as suggested by the induced magnetic field (Kivelson et al., 2002), it should be located between the ice-I and the ice-V layers (**Figures 6** and **8**). If the liquid layer is pure water, and if there is no other internal heating source than the radiogenic decay in the silicates (0.6 TW, **Table 2**), then the scaling laws describing stationary convection suggest that the ice viscosity should have a value around 10^{15} Pa s (**Figure 12**). Such a value is much larger than the viscosity of ice measured on terrestrial glaciers. If the melting-point viscosity is comparable with that of terrestrial glaciers (5×10^{13} Pa s), then the amount of heat that can be removed is on the order of 1.5 TW. Another source of heat, for example, tidal heating, is required to explain such a large value. According to simulations (Showman and Malhotra, 1997; Sotin and Tobie, 2004; Tobie et al., 2005), present-day tidal heating cannot be larger than radiogenic heating and cannot explain such a large value. Another explanation invokes the presence of antifreeze materials that would lower the temperature (increase the viscosity) at the solid–liquid interface (Spohn and Schubert, 2003). But this is clearly not proven.

The differentiation of the metallic (largely iron, based on solar abundances) core may have been a late episode in the evolution of the icy satellites (Deschamps and Sotin, 2001; Grasset and Sotin, 1996). This hypothesis presumes that the outer ice layer acts as an insulator that cannot release all the heat produced by the decay of the radiogenic elements. How the internal dynamics and evolution scenarios, including core formation, are related to Ganymede's geology – characterized by two types of terrain, the heavily cratered dark terrain and the less cratered bright terrain (e.g., Patterson et al., 2010) – is still an open question. Images obtained by *Voyager* and *Galileo* suggest that Ganymede experienced some resurfacing events at about 3.5 Ga. In the case of Ganymede, the link between orbital evolution into the Laplace resonance and thermal evolution by heat pulses through enhanced eccentricities is a promising scenario (Showman and Malhotra, 1997). As suggested by Bland et al. (2009), this may also explain phases of global expansion that occurred during Ganymede's history as derived from extensional features on the surface.

Callisto's radius is about 200 km smaller than that of Ganymede and its mass is 70% that of Ganymede. The old, heavily cratered, and dirt-rich surface suggests that endogenic resurfacing has never happened since the satellite completed its accretion. The post-*Galileo* view of Callisto is surprising. If hydrostatic equilibrium is achieved, the *Galileo* gravity data suggest that the moment of inertia is equal to 0.3549 ± 0.0042 (Schubert et al., 2004). This value is not compatible with a fully differentiated Callisto (Anderson et al., 2001b). It should be noted, however, that the *Galileo* flybys were near equatorial, implying that the gravity field coefficients J_2 and C_{22} could not be determined independently. Furthermore, the derived S_{22} value is nonzero, suggesting deviations from the hydrostatic state. In the case of Callisto, relatively large uncertainties about the internal structure and hydrostaticity remain that could best be addressed by tracking polar (or highly inclined) flybys and orbits of future missions complementing the *Galileo* data.

The magnetic data suggest that an ocean is present at depth around 150 km (Khurana et al., 1998; Zimmer et al., 2000). These two interpretations of geophysical data seem contradictory since the presence of an ocean would probably lead to differentiation. In order to reconcile these two observations,

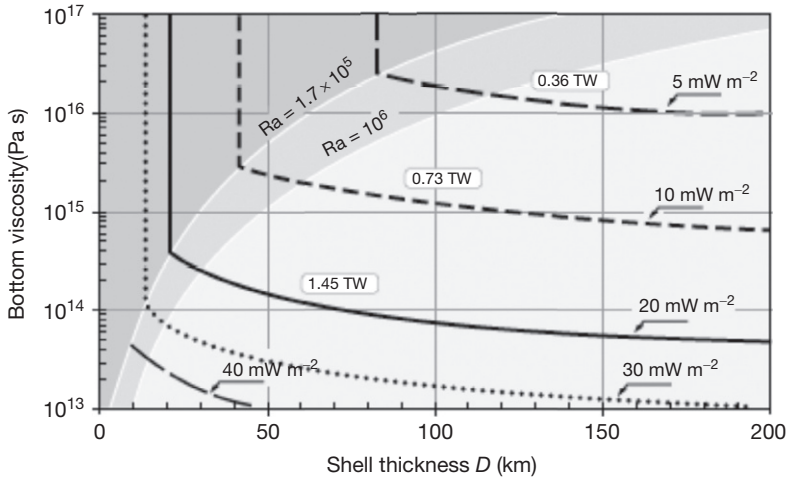


Figure 12 Amount of heat transferred by conduction or by convection within the ice-I shell of a large icy satellite, plotted as bottom viscosity, required to match a given heat flux versus thickness of the shell. Lines of constant total power and surface heat flow (power per unit area) are shown. Parameters for Callisto have been used.

Nagel et al. (2004) propose that Callisto's core is a mixture of ice and rock with the rock concentration near the close-packing limit.

Models of heat transfer in the ice-I outer layer of Callisto vary in their results. One model suggests that convection is not possible, which explains the presence of the ocean (Ruiz, 2001). McKinnon (2006) suggests that the cooling from above of the ice-I shell leads to subsolidus convection. This latter work confirms the results of previous modeling studies (e.g., Deschamps and Sotin, 2001). The amount of heat that can be removed by subsolidus convection is the same whether one uses the scaling laws proposed by Dumoulin et al. (1999) or those described in Deschamps and Sotin (2001). The amount of heat transferred depends strongly on the viscosity and weakly on the thickness (**Figure 12**). This amount is much larger than the heat released by the decay of the radiogenic elements at present time (0.4 TW, **Table 2**) unless the viscosity of ice near its melting point is larger than 10^{15} Pa s. Convection is able not only to remove the radiogenic heat but also to freeze the ocean very quickly unless some antifreeze contaminants like ammonia are present (Deschamps and Sotin, 2001; McKinnon, 2006; Spohn and Schubert, 2003).

The comparison between Ganymede and Callisto is interesting since these two satellites have mass and radius that are fairly similar but their internal structure is apparently quite different. Ganymede is fully differentiated with a liquid iron core generating a magnetic dynamo, whereas Callisto is only partially differentiated. Another difference is their surfaces. Ganymede's surface was tectonically active, whereas Callisto's surface does not show any sign of late activity. More information about their interiors is needed to understand why they are so different. The lower radiogenic power within Callisto may be one explanation. Alternatively, different accretion scenarios, due to different locations in the Jovian subnebula (Barr and Canup, 2008; Canup and Ward, 2002; see also **Chapter 10.17**), may be considered. This would be closely connected with the evolution of Ganymede into the Laplace resonance locking. Callisto has not been captured in resonance with the other satellites, and tidal heating never has played a significant role in its evolution. Both a primordial origin (Peale and Lee, 2002) and an evolution into resonance at a later stage (Bland et al., 2008; Showman and Malhotra, 1997) may have caused a difference between the two satellites with respect to internal structure and geology. Barr and Canup (2010) suggest different impact rates and thus higher amounts of energy provided to Ganymede's surface and interior (as compared to Callisto) during late heavy bombardment as a driving factor for the different evolutionary paths.

The example of Ganymede and Callisto shows that not only the bulk composition but also the interaction with other satellites or with the subnebula during the phase of accretion determines the satellite evolution. Evolution scenarios of icy moons in many cases cannot be considered separated from each other but must be seen in the context of the whole satellite system.

10.18.7.2 Europa

With respect to the distance from Jupiter, Europa is the second of the Galilean moons (note that there are additionally four small satellites, Metis, Adrastea, Amalthea, and Thebe inside Io's orbit, which is the innermost of the Galilean satellites), orbiting around the giant planet at a period of 3.55 days. Its density of 3018 kg m^{-3} lies between that of Io (3530 kg m^{-3}) and those of Ganymede (1936 kg m^{-3}) and Callisto (1851 kg m^{-3}). The mass fraction of ice on Europa is only about 6–9% based on the mean density. Thus, Europa is a fundamentally rocky body with only small amounts of H_2O. The characteristics of the gravity field have been measured during several Europa flybys in the course of the *Galileo* mission (Anderson et al., 1998; Schubert et al., 2004). Unfortunately, the flybys were not with a geometry that allowed J_2 and C_{22} to be determined independently from each other. Assuming hydrostatic equilibrium ($C_{22} = 0.3\ J_2$), it is possible to determine the value of the hydrostatic Love number k_f, which is related to the moment of inertia C about the rotation axis by the Radau formula (eqn [3]). The value of 1.048 ± 0.020 found for Europa suggests that it is fully differentiated including an iron-rich core, a silicate mantle, and an H_2O shell (Anderson et al, 1998; Schubert et al., 2004, 2009; Sohl et al, 2002). Another way to interpret the density data, which cannot be completely ruled out on the basis of the MoI value, is that Europa would be made of a large core not separated in metal and rock overlaid by a layer of H_2O (e.g., Schubert et al., 2009). However, such a structure is unlikely because the required mean density of the rock/iron mixture would exceed the mean density of Io. Enrichment in heavy elements in Europa would, however, be difficult to understand for a satellite smaller than Io that has formed further away from Jupiter in a colder region of the Jovian subnebula. Typical density models of Europa assume three-layer structures with core sizes ranging roughly between 300 and 700 km and thicknesses of the H_2O layer between 80 and 170 km. The values depend on the assumed core and silicate densities (Anderson et al, 1998; Schubert et al., 2009; Sohl et al., 2002). The core density would be significantly reduced by volatiles, for example, sulfur. The mantle density is likely to range between 3000 and 3500 kg m^{-3} constrained by Io's density (upper bound) and likely lacks hydrated silicates (lower bound). Under given pressure conditions on Europa, hydrated rock would release its water at temperatures as low as 500–800 °C (Schubert et al., 2004), which are difficult to avoid in Europa's evolution, especially in the view of tidal heating and differentiation processes that have occurred on Europa.

Additional information about the internal structure of Europa comes from magnetometer measurements during the *Galileo* mission. Europa has a strong inductive magnetic field (see also **Section 10.18.5.5**) consistent with an electrically conducting interior with an outer radius equal to nearly the entire radius of the moon (Khurana et al., 1998; Zimmer et al., 2000). However, it does not show any evidence of a permanent dipole. The induction field can be explained by a water ocean of similar conductivity to the Earth's oceans, provided this ocean has a thickness exceeding 10 km (Kivelson et al., 2000). These measurements of an induced magnetic field provide the strongest observational evidence for the present-day existence of an ocean.

The high-resolution images obtained during the *Galileo* mission provide some clues about Europa's internal structure and evolution. Compared to the other icy Galilean satellites,

Europa's surface displays intriguing tectonic features, such as domes, disrupted blocks, chaotic terrains, and fractures (Carr et al., 1998; Collins and Nimmo, 2009; Collins et al., 2000; Doggett et al., 2009; Greeley et al., 1998, 2004; Greenberg et al., 1998; Kattenhorn and Hurford, 2009; Pappalardo et al., 1998, 1999; Prockter and Patterson, 2009; Riley et al., 2000; Spaun et al., 1998). The small number of impact craters suggests that Europa's surface has undergone substantial resurfacing by either tectonic or cryovolcanic processes. It is younger than the surfaces of the two other icy Galilean satellites Ganymede and Callisto. However, it is difficult to obtain absolute ages of surfaces from the number of impact craters, because the cratering rate in the outer solar system is not well known. Models suggest for the surface of Europa ages between 10 million years and several hundred million years (Bierhaus et al., 2009 and references therein). Another possibility for assessing the age is the erosion rates by sputtering. The estimates of erosion rates by sputtering are constrained by the measurements of the H_2O escape rate during the *Galileo* mission. If sputtering has been operating for more than 1 billion years, it should have erased most of the topography; however, this is not the case for Europa. The remaining topography of Europa is consistent with an age of its surface in the range of some million years (Pappalardo et al., 1999).

The disrupted blocks, also called 'rafts,' have been interpreted by different models. Collins and Nimmo (2009) list the following models for formation of chaotic terrain: (a) melt through, (b) diapirism, (c) brine mobilization by diapirism, (d) brine mobilization driven by partial melt through, (e) sill formation, and (f) impacts. The melt-through and brine mobilization models can explain key features of chaos terrain. However, it is still not clear how both large ($\sim$1000 km) and small ($\sim$1 km) features can be generated by these processes.

In the thin-shell model, the outer solid layer is less than a few kilometers thick and the surface is constantly in contact with an ocean below due to tidal cracking and disruption. Small variations in the thermal gradient cause chaos and lenticulas (Carr et al., 1998; Greenberg et al., 1999). In another thin-shell model (Thomson and Delaney, 2001), the thin crust is periodically broken by faults that form in response to the tidal forces resulting from Europa's eccentric orbit around Jupiter. The blocks resulting from the faulting of the ice crust could then be entrained by the streams of the liquid ocean and could move relative to each other. Because the surface temperature is far below the freezing point, the displacement of the blocks would be limited by the rapid freezing of any liquid that would be present in between the blocks.

Schmidt et al. (2011) suggest that chaos terrain forms above liquid water lenses perched within the ice shell. By their model, topographic heights of chaos terrain, for example, *Conamara Chaos*, can be explained by refreezing of liquid reservoirs. Active melting would result in topographic depressions, as observed for *Thera Macula* a region potentially showing active resurfacing.

Another reason for the crust to be disrupted is the possibility for localized sources of heat in Europa's silicate layer (e.g., Goodman and Lenferink, 2012). It may be conceivable that tidal forces would lead to partial melting in Europa's silicate layer (e.g., Moore and Hussmann, 2009) and volcanism at the silicate–water interface. Such a scenario is inspired by Io's very active volcanism. In this case, the water would be very hot during a volcanic eruption and would rise to the surface, melting the ice crust and inducing strong displacements. However, the amount of heat dissipated in Europa's silicate shell and deep interior is not well constrained and whether it would allow for silicate volcanism is presently unknown.

It must be noted that volcanism is not necessarily required. Convection in Europa's silicate layer would create zones of higher temperature, and percolation of water within the silicates would create conditions favorable for hydrothermalism (Lowell and DuBose, 2005) and subsequent displacement of hot water to the surface. Such a scenario is appealing for exobiology because the conditions at the silicate–ocean interface would be very similar to those existing on Earth's seafloor.

An alternative model suggests that the blocks and domes are related to subsolidus (solid-state) convection in the lower part of the crust (McKinnon, 1999; Nimmo and Manga, 2002; Pappalardo and Barr, 2004; Pappalardo et al., 1998; Rathbun et al., 1998; Sotin et al., 2002). In this model, the domes are explained by the topography resulting from the rising of hot subsolidus icy plumes forming at the interface between the liquid layer and the icy shell. The topography of the domes is not well constrained and cannot be used to constrain the characteristics of the upwelling plumes. If this is the right explanation for the domes, it implies that the icy crust is at least 20 km thick. The chaotic terrain would form by the collapse of the upper crust as depressions form due to partial melt in the plume head induced by tidal heating (Sotin et al., 2002). Convection models that include surface weakening and compositional buoyancy are promising approaches in explaining some chaos terrains, pits, and uplifts (Barr and Showman, 2009). However, the current data are likely to be too limited to unambiguously determine the formation processes of the different features.

Europa is unique in displaying large-scale cycloidal fractures on its surface. The formation of these features is closely connected to the diurnal tides on Europa (Hoppa et al., 1999). As tidal stresses build up during one orbital cycle, cracking occurs when the tensile strength of the ice is reached. The opening crack would then propagate according to the time-dependent stress field dictated by the tidal potential. When the tensile stress becomes insufficient to continue crack propagation, it would stop at a certain point. After a few hours when the tensile stress has built up again, the crack would continue to propagate. However, the direction would be different because of the different tidal stress field. The resulting cracks form arcuate segments connected to a cycloidal chain with lengths of several hundreds or thousands of kilometers. The formation of these features requires sufficient tidal amplitudes that are only occurring when a thick fluid layer is present below the ice shell. The cycloidal features therefore provide strong evidence for a subsurface ocean on Europa at the time they formed. Further evidence for such a global decoupling layer comes from *Galileo* imaging analysis, suggesting nonsynchronous rotation of the ice shell on a timescale of $\sim$10 000 years (Geissler et al., 1998). By the analysis of orientation and distribution of lineaments on Europa, it was found that the ice shell rotates slightly faster than synchronous. Based on gravity data, it is inferred that the bulk of Europa is

rotating synchronous. Therefore, nonsynchronous rotation would imply a decoupling layer between the ice shell and the deep interior.

Albeit on a smaller level as compared to Io, Europa is subject to tidal heating (Cassen et al., 1979, 1980; Hussmann et al., 2002; Ojakangas and Stevenson, 1989; Ross and Schubert, 1987; Sotin et al., 2009; Squyres et al., 1983b; Tobie et al., 2003, 2005). The heat generated by tidal flexing can be more than one order of magnitude larger than radiogenic heating alone and could explain the presence of an ocean below the icy crust. Since the amount of ice is small, high-pressure phases cannot form and the ocean would be in contact with the silicate core, which makes Europa quite appealing from an astrobiology point of view.

There are also a few multiringed structures on Europa, such as Tyre and Callanish, which are interpreted as impact features. Studies of the morphology of impact craters suggest an ice thickness at the time and location of the impact of about 4–20 km, converging on the range of 15–19 km (Schenk and Turtle, 2009). This interpretation of the formation and relaxation of impact craters is a strong argument in favor of a deep liquid ocean below an icy crust. Although an entirely solid convecting layer of 'warm ice' may also be consistent with crater relaxation, such a structure is unlikely because a thick low-viscosity layer would lead to vigorous tidal heating exceeding by far the heat loss due to convection. The disequilibrium between heat production and heat flow would cause melting and thinning of the ice layer that would again imply an ocean layer at the base of the ice shell (Hussmann et al., 2002).

With the progress made in modeling thermal convection in planetary interiors, models describing Europa's thermal evolution have substantially improved in the last 20 years. The main important improvements are the description of convection with complex temperature and compositional dependencies (e.g., Barr and Showman, 2009 and references therein) and the coupling between orbital characteristics and tidal heating, originally proposed for Io (Fischer and Spohn, 1990; Greenberg, 1982; Yoder and Peale, 1981) and later applied to Europa (Hussmann and Spohn, 2004; Tobie et al., 2005).

As stated before, the amount of tidal heating is much larger for Europa than for the two Galilean satellites Ganymede and Callisto or for Saturn's moon Titan. It may be more than an order of magnitude larger than radiogenic heating. Tidal dissipation is mainly located in the convective layer of the ice shell, that is, in the region where the temperature is close to the melting point (see **Figure 11**). The very effective tidal heating within a thick high-temperature layer would prevent the ice layer from freezing because with increasing ice thickness, the heating rate will also increase significantly, exceeding the rate of heat transfer. The equilibrium thickness in which the tidal heating rate in the ice shell and the radiogenic heating from the rocky mantle are balanced by the convective heat flow is of the order of 30–40 km (Hussmann et al., 2002), which is less than the total thickness of 70–200 km of the total H_2O layer derived from the gravity data, therefore suggesting the presence of an ocean.

Tidal heating in a convective shell explains the presence of an ocean within Europa even without invoking the presence of antifreeze material such as ammonia. However, it does not mean that antifreezing materials are absent. The induced magnetic field indicates that the ocean contains ions, as electrical conductors, and therefore, a reduction of the melting point can be expected. Inferences about the composition and evolution of Europa's ice–ocean–rock system mainly rely on modeling and laboratory data (Zolotov and Kargel, 2009). Data from future space missions, in particular spectroscopic observations of the surface and the search for active plumes, would be required to further assess possible ocean composition and its astrobiological potential.

Consistent with geologic observations (e.g., Figueredo and Greeley, 2004), the thickness of Europa's ice shell additionally may be subject to considerable changes with time, adjusting to the corresponding heat budget (e.g., Nimmo and Manga, 2009). The latter varies not only due to the decay of radioactive isotopes, which would imply an increase of thickness due to the decreasing heat source, but also due to the orbital state and the resulting tidal heating. The tidal deformation and also the tidal dissipation rate will depend on the orbital eccentricity and mean motion (eqn [10]). Since Europa is locked in resonance with Io and Ganymede, its eccentricity is forced and oscillations in both the orbital state and the heat production rate may occur as a response to Io's thermal–orbital state, which is mainly controlled by tidal interaction between Io and Jupiter (Yoder, 1979; Yoder and Peale, 1981). Oscillatory phases have been proposed for Io by Ojakangas and Stevenson (1986) and by Fischer and Spohn (1990). As a consequence of such a thermal–orbital coupling, the thickness of Europa's ice shell may vary between a few kilometers and a few tens of kilometers on a timescale of the order of 100 Ma (Hussmann and Spohn, 2004). The observed diversity of geologic surface features on Europa may be partly explained by these time-dependent variations of ice thickness.

Because of the resonance locking, Europa's eccentricity is maintained on geologic timescales. Therefore, tidal heating would not cease, which is essential for any biological evolution that might have started at some point in Europa's history. Depending mainly on the tidal interaction between Jupiter and Io, tidal heating may operate on Europa for Gy preventing the ocean from freezing and adding heat to Europa's interior. Europa is therefore a very attractive body for exobiology (Greenberg et al., 2000; Hand et al., 2009; Reynolds et al., 1983), even more because the water layer is likely to be in contact with the silicate core, allowing for complex chemical processes (Kargel et al., 2000; Zolotov and Kargel, 2009). At this interface, hydrothermal systems similar to those at terrestrial seafloors may play a role (Lowell and DuBose, 2005). It has been envisaged that the conditions may be analogous to those existing at terrestrial black smokers where life develops without sunlight and where specific bacteria form. However, there is no evidence, yet, that Europa's silicate core is subject to volcanism and that tidal heating in the silicate shell is capable of producing zones of high heat flux as in the case of Io.

Recent numerical evolution models of Europa are also focusing on the dynamics within the ocean and its implications for heat generation and heat transport (Tyler, 2008; Chen et al., 2013) and on the salinity within the ocean (Travis et al., 2012; Vance and Goodman, 2009). Theoretic models of the rotation state of Europa (including small obliquities) have been investigated to infer possibilities to constrain the ice thickness above a decoupling ocean layer by measuring the libration amplitudes (Van Hoolst, 2008; Jara-Orue and

Vermeersen, 2014) and to understand the interplay between solid body-tides, the decoupling mechanisms, and the physical librations (Van Hoolst et al., 2013).

10.18.7.3 Titan

Titan, the largest satellite of Saturn, is unique among the moons in the solar system because of its substantial atmosphere and photochemically produced, dark orange stratospheric haze layer that precludes direct observations of the surface. Detailed information on the surface, however, was obtained from the landing of the *Huygens* probe (Lebreton et al., 2005) and by *Cassini* RADAR data collected over the course of the mission (Elachi et al., 2005; Stofan et al., 2006; Lopes et al., 2010). Titan's dense atmosphere that is dominated by molecular nitrogen with a substantial component of methane (Coustenis et al., 1989; Flasar et al., 2005; Owen, 1982) suggests that the satellite's interior is more enriched in volatile ices such as $NH_3 \bullet H_2O$ and $CH_4 \bullet nH_2O$, which have densities similar to that of water ice (Hunten et al., 1984). This is likely to be due to the lower temperature of the less massive proto-Saturnian nebula in comparison with the proto-Jovian nebula, which favors the incorporation of a larger amount of volatiles such as ammonia and methane at the distance of Titan (20.6 Saturn radii) without significantly affecting the satellite's mean density in comparison with Ganymede and Callisto.

The absence of rocky inner satellites in the Saturnian system may indicate lower temperatures of the nebula during the evolution of the gaseous and dusty accretion disk from which the satellite system formed by mutual collisions of icy and rocky planetesimals (Coradini et al., 1995; Owen, 2000). Based on the assumption that all nitrogen was originally supplied as NH_3, volatile components could appear in roughly cosmic abundances in the interior of Titan with ratios of $NH_3/H_2O \approx 1/6$ and $NH_3/CH_4 \approx 1$. Corresponding water ice–ammonia mixtures then have an ammonia content of 14 wt%, a composition more water-rich than the peritectic composition of about 32 wt% (Hunten et al., 1984; Lewis, 1972; Stevenson, 1992).

The presence and abundance of atmospheric methane that would be lost over the age of the solar system due to photolytic dissociation (Strobel, 1974) suggests that sources of methane must be present at the surface and possibly in Titan's interior (Sotin et al., 2005; Tobie et al., 2006). Such a connection between interior, surface, and the dense atmosphere allowing for Titan's methane cycle (Atreya et al., 2006) is unique for the icy satellites.

Titan is intermediate between the Jovian satellites Ganymede and Callisto with respect to its radius of 2575 km and mean density of 1881 kg m^{-3}. The mean densities of Ganymede, Callisto, and Titan indicate that their interiors are composed of ice and silicates in nearly equal amounts by mass (Lupo, 1982; see in the preceding text). The interior of Titan is likely to be differentiated at least into a rock + iron core and an icy mantle as a consequence of substantial accretional heating accompanied by partial outgassing and atmosphere formation (e.g., Grasset et al., 2000; Hunten et al., 1984). Based on *Cassini* flybys, Iess et al. (2010) have derived a moment of inertia of 0.3414, which is too high for fully differentiated models as obtained, for example, for Ganymede. Their interpretation is either an incomplete separation of rock from ice or a core in which a large amount of water remains chemically bound in silicates. Barr et al. (2010) suggested that a partial differentiation of Titan arises from undifferentiated accretion and impact-induced core formation permitting some of Titan's rock to consolidate into a core. Gao and Stevenson (2012) had argued that the actual moment of inertia may be less than the reported value because of possible nonhydrostatic effects that render invalid approximations used in Iess et al. (2010). If the maximum value of the MoI is used, then Titan has either a mixed rock–ice layer or a substantially hydrated silicate core (Castillo-Rogez and Lunine, 2010). The former is difficult to maintain (O'Rourke and Stevenson, 2013), while the latter requires substantial early leaching of radioactive isotopes into the overlying ocean (see the succeeding text) to avoid complete dehydration.

The orbital eccentricity of Titan causes tidal variations of the quadrupole moments from which the second-degree potential Love number k_2 has been derived (Iess et al., 2012). k_2 is related to the satellite's interior structure and rheological properties and reflects the gravitational signature of a substantial internal ocean. Indeed, the value of k_2 seems to be between 0.6 and 0.7 with the expected value of 0.45–0.5 about 2σ away from the mean. If further gravity flybys confirm the high value, then either Titan has solid layers that are highly compressible in addition to the ocean, or the ocean is substantially overdense compared to pure water, perhaps containing substantial amounts of sulfur as in Fortes (2012).

Evidence for subsurface oceans from induced magnetic fields as in the case of the Galilean moons could not be obtained in the Saturn system. The main reason is that the magnetic field axis of Saturn is not tilted significantly with respect to the rotational axis as it is the case for Jupiter. Therefore, variations in the field strength at the location of the moons are much smaller and could not be detected. Furthermore, Titan's orbit is relatively far away from Saturn. However, there is further evidence for a subsurface ocean on Titan based on the detection of the second harmonic of a Schumann resonance (Beghin et al., 2012) during the descent of the *Huygens* probe. The electric current sheets induced in Titan's ionosphere by Saturn's magnetospheric plasma flow could trigger electromagnetic waves in Titan's atmosphere that would be trapped in a Schumann resonance between two electrically conducting layers. The inner reflecting boundary is proposed to be a buried water–ammonia ocean lying at a likely depth of 55–80 km below a dielectric icy crust. The almost perfect electrical conductivity in the subsurface could be provided by such a global water ocean.

Further evidence for a subsurface ocean on Titan could be obtained by precisely determining the rotational state of the satellite. Current estimates of the moment of inertia from Titan's spin-pole orientation that are independent from the MoI determination from the gravity field differ from the latter. This discrepancy can be explained by a decoupling of the ice shell from Titan's deep interior, presumably by a subsurface ocean. However, the spin-pole orientation may also be affected by atmospheric excitations. To distinguish between these two models may be possible from further data analysis of *Cassini*'s extended mission (Bills and Nimmo, 2011).

Depending on the amount of volatiles incorporated into the icy mantle and depending on the satellite's thermal evolution, an ammonia-rich liquid water layer may be located between the near-surface ice-I layer and solid high-pressure ice layers at greater depth (Grasset and Sotin, 1996; Grasset et al., 2000; Sohl et al., 2003).

Because of Titan's close resemblance to Ganymede in mass and overall density, the evolution of the two bodies should be similar assuming identical compositions and boundary conditions. However, a number of arguments have been advanced that Titan incorporated some amount of ammonia in its interior during formation (e.g., Niemann et al., 2005) and might have also incorporated large amounts of methane (Tobie et al., 2006). The former implies a substantial liquid layer would exist throughout Titan's history, and the latter implies significant changes in the thermal properties of the upper ice-I crust if the methane were to force its conversion to methane clathrate hydrate (Lewis, 1971).

A pristine water-rich ammonia–water ocean may have formed as a consequence of the blanketing effect of a hot protoatmosphere, if accretion was completed within 10^4–10^5 years (Kuramoto and Matsui, 1994). This would freeze from top and bottom. It has been suggested that the ice-I layer on top of the ocean was initially cooling and thickening due to thermal conduction and became unstable to convection about 200 My after accretion (Grasset and Sotin, 1996). Some volatiles will presumably be present in the form of hydrates and clathrates, thereby affecting the rheological and thermal properties of the ice layers (Hunten et al., 1984; Stevenson, 1992; Tobie et al., 2006). The presence of volatiles tends to reduce the freezing temperature of the ocean, thereby lowering the temperature of the ice-I layer and increasing its viscosity. Larger viscosities due to temperatures far below the melting temperature then would retard convection and slow cooling of the ice-I layer, so that the subsurface ocean may have survived to the present day because of the progressing enrichment of ammonia as compared to the primordial ocean (Deschamps and Sotin, 2001; Grasset and Sotin, 1996).

Furthermore, during Ganymede's evolution into the Laplace resonance with Europa and Io, it may have been strongly heated, which possibly led to complete differentiation as described in the subsection earlier. Titan almost certainly experienced no such event though it is subject to periodically varying tidal distortions resulting from its significant orbital eccentricity of 0.0292 in addition to its permanent rotational flattening.

Evolution models instead are constrained to avoid excessive dissipation in Titan's interior throughout its history so as to maintain its observed present-day free eccentricity (Sohl et al., 1995; Tobie et al., 2005). There is considerable uncertainty in the initial eccentricity, both generally and specifically because Titan might have suffered collisions early on from one or more fragments from the breakup of proto-Hyperion (Farinella et al., 1997).

Models in which Titan is volatile-rich (ammonia and methane) evolve with time so as to maintain a very thin ice-I crust, typically 20 km or less, up to the last half billion years of Titan's history, because of the low thermal conductivity and high rigidity of methane clathrate hydrate in contrast to ice-I (Tobie et al., 2006). This, along with the onset of convection in the rocky core halfway through Titan's history and the gradual thickening of the crust via freezing of ice-I in the last half billion years, provides a potential setting for repeated outgassing of methane to the surface to replace that destroyed by atmospheric photo- and charged-particle chemistry (Lorenz and Lunine, 2005). However, the lines of evidence for the early presence of ammonia and outgassing of methane through geologic time are indirect, principally through the abundances of nonradiogenic and radiogenic argon (Niemann et al., 2005), and so the models must be regarded with caution.

10.18.7.4 Enceladus

As revealed by the *Cassini* mission, Enceladus is one of the very few planetary bodies in the solar system for which present-day activity was confirmed (Porco et al., 2006). Considering only the terrestrial planets and the satellites (not including comets, for instance), Enceladus is by far the smallest among these objects, with the others being the Earth, Io, and Triton. The huge amount of energy required to drive the cryovolcanic plumes, which are the dominant source of Saturn's E ring (Kempf et al., 2010; Spahn et al., 2006), and to cause the high temperatures and heat flow detected near Enceladus' south pole (Howett et al., 2011; Porco et al., 2006; Spencer and Nimmo, 2013; Spencer et al., 2006, 2009) still longs for an explanation. The geologically young terrain at several places on Enceladus and the recently strongly modified surface in the south-pole region further suggest that Enceladus contains internal heat sources. The plumes and the maximum temperatures correlate with the so-called tiger stripes, parallel fractures at the south pole, approximately 130 km long and separated by ~35 km (Porco et al., 2006; Spencer et al., 2006). The tiger stripes have raised margins enclosing a central trough approximately 2 km wide and 0.5 km deep (Spencer and Nimmo, 2013). From high-resolution imaging, the sources of erupting jets have been identified as the warmest regions within individual tiger stripes (Spitale and Porco, 2007). The strong correlation between the brightness of the plume and the orbital phase (Hedman et al., 2013) demonstrates the role of tidal forces in controlling the dynamics of the jets that compose the plume: at apoaxis, tidal forces create extensive stresses along the tiger stripes allowing material to be extruded more easily than at apoaxis (Hurford et al., 2007).

The nature of the heat-generating processes still remains unclear. One suggestion, already discussed in earlier works and motivated by surface images from the *Voyager* flybys, is tidal heating (Peale, 2003; Poirier et al., 1983; Ross and Schubert, 1989; Squyres et al., 1983a). Due to its proximity to Saturn and its elliptical orbit, Enceladus is subject to periodic tidal deformation, which will lead to internal friction and thus heating. Enceladus' eccentricity is forced by Dione, which is locked in a 2:1 mean motion resonance with Enceladus, thus keeping the eccentricity finite, presently at 0.0045. Depending on the internal temperature and rheological state of the satellite, tidal friction may be sufficient to power the observed activity. From *Cassini*'s composite infrared spectrometer instrument, an output of at least 5.8 ± 1.9 GW at the south-pole hot spot was derived (Spencer et al., 2006, 2009), which would be within the range of heat calculated from global tidal heating models under certain conditions. More recent estimates (Howett et al., 2011) suggest a heat output of 15.8 ± 3.1 GW. However, due to the complexity of the measurement, the true uncertainty might be larger (Spencer and Nimmo, 2013). A liquid layer decoupling the ice shell from the rock core would additionally enhance the tidal heating rate (as compared to a completely solid body) by approximately an order of magnitude. Global tidal heating models, however, cannot explain the asymmetry between the

north pole and the south pole of Enceladus by itself. Furthermore, Roberts and Nimmo (2008) had shown that under present orbital conditions, the heat flow would exceed the heat production rate due to tides and radiogenic heating, leading to freezing of the entire ocean. To prevent the ocean from freezing, specific assumptions on the rheology of the ice, on the orbital evolution, and/or on the composition have to be made.

Pockets of liquid water and local tidal heating restricted to the south-pole region is another option (Behounkova et al., 2012; Collins and Goodman, 2007; Tobie et al., 2008). In all these models, an additional process triggering vigorous tidal heating is required. A runaway process, similar to the one suggested for the rocky mantle of Io by Peale et al. (1979), is problematic because it would be expected rather for the innermost satellite Mimas, than for the more distant Enceladus. Under similar rheological conditions, tidal heating rates are at least ten times greater on Mimas, due to its proximity to Saturn and its greater orbital eccentricity (Squyres et al., 1983a). The striking differences between the two satellites, with Mimas being heavily cratered and geologically inactive for billions of years and Enceladus showing an enormous internal activity, are thus even more puzzling. Schubert et al. (2007) suggest that the higher rock content of about 60%, which was derived from the determination of the mean density by *Cassini* (Porco et al., 2006) compared to 20% in Mimas and the resulting higher radiogenic heating rates in Enceladus, can lead to the obviously different evolutionary paths.

Meyer and Wisdom (2007a) computed the tidal heating rate based on the assumption that the Enceladus–Dione resonance is currently in equilibrium. This would imply that the rate of decrease in orbital eccentricity due to dissipation in Enceladus is compensated by the rate of increase in eccentricity due to tidal torques by Saturn and the resulting increased eccentricity forcing by the resonance with Dione. Independent from a specific rheological model of Enceladus, the equilibrium heating rate would not be sufficient to explain the present-day heat flow of Enceladus (Meyer and Wisdom, 2007a). It should be noted that the calculated heating rate does depend on the assumed quality factor Q_S of Saturn. $1/Q_S$ is a measure for the energy dissipated in Saturn during one orbital cycle. It thus determines the total amount of energy available for the Enceladus–Dione system. The greater Q_S, the lesser dissipative is Saturn. However, for reasonable values of Q_S, the models do not match the measured heat flow values (e.g., $Q_S = 18\,000$ suggested as a lower bound by Meyer and Wisdom (2007a) would imply a maximum equilibrium heating rate in Enceladus of 1.1 GW).

There is evidence from the relaxation of craters (Bland et al., 2012) and the analysis of flexural uplift along a rift zone (Giese et al., 2008b) that Enceladus has passed through several periods of enhanced heating in its past. It should be noted that these high-heat-flow regions are not confined to southern latitudes only. Analogous to Io in the Jupiter system, episodic heating including periodic changes in eccentricity and heat production (Ojakangas and Stevenson, 1986) might be a possibility. However, according to Meyer and Wisdom (2008), the application of this model does not lead to oscillations about the tidal equilibrium in the case of Enceladus. A detailed analysis of the coupled thermal–orbital evolution including temperature-dependent rheological parameters and structures of Enceladus within the resonance with Dione would be required to understand Enceladus' history.

Possible modifications of the tidal scenario include a non-smooth evolution into the 2:1 Enceladus–Dione mean motion resonance involving large eccentricities, a past resonance locking with other satellites including the small satellite Janus enhancing the forced eccentricity of Enceladus (Lissauer et al., 1984), or possible spin–orbit secondary resonances before Enceladus reached its synchronous rotational state (Wisdom, 2004). The first alternative is problematic because the evolution into resonance could also be achieved smoothly, without involving chaotic behavior and large eccentricities (Peale, 2003). Meyer and Wisdom (2007b) have investigated Enceladus' evolution into the current resonance with Dione including several resonance passages. These simulations suggest that the Enceladus–Dione 2:1 resonance is entered close to tidal equilibrium, implying heating rates around 1.1 GW, only. However, a detailed dynamic evolution of Enceladus into the resonance including tidal heating and thermal–orbital coupling, that is, time-dependent rheological parameters, has not yet been undertaken. The third alternative would enhance the tidal heating rate by two or even three orders of magnitude if Enceladus was locked in a 3:1 secondary spin–orbit resonance (Wisdom, 2004) and could thus easily account for the difference between Enceladus and Mimas for which such a coupling cannot be expected. However, for the *Cassini*-derived shape parameters, these spin–orbit couplings are unlikely to be obtained by Enceladus.

If the physical librations of Enceladus' ice shell are large enough, they might cause significant tidal heating in the ice shell (Hurford et al., 2009). However, large libration amplitudes could not yet be confirmed by *Cassini*. The libration amplitude would be significantly enhanced if the ice shell is decoupled from the deep interior by a global ocean (as, e.g., suggested by Patthoff and Kattenhorn, 2011). Therefore, detecting the diurnal libration and measuring its amplitude is a crucial experiment to distinguish between global and local ocean models.

All the models involving global tidal heating cannot explain the strong asymmetry between the two poles of Enceladus, and to concentrate the heat sources within one relatively small region still is an open issue. Suggested mechanisms to enhance tidal heating locally include one-degree convection (Grott et al., 2007) and a locally mechanically weak lithosphere (Barr, 2008; Han et al., 2012; O'Neill and Nimmo, 2010; Roberts and Nimmo, 2008). Other processes, for example, shear heating (Nimmo et al., 2007), may be triggered by such hemispheric or local asymmetries. Reorientation would later drive the active region toward the south pole (Nimmo and Pappalardo, 2006).

As mentioned earlier, a liquid layer would enhance tidal heating significantly. However, radiogenic heating due to long-lived isotopes will be insufficient to maintain a liquid water shell up to the present time. Enceladus' mean density of 1609 kg m^{-3} (Thomas, 2010) suggests a rock-to-ice ratio of about 60–40 wt%. Although it is very likely, with regard to the energetic processes in its interior, that Enceladus is differentiated into a rock core of about 150–170 km thickness and an H_2O layer, there is no such evidence from independent measurements. Gravity flybys to determine J_2 and C_{22} are

still being analyzed (as of August 2013), and also the shape data showing several regional depressions (Schenk and McKinnon, 2009) are not easy to interpret. The differences of the satellite's main axes can be used to verify whether it has assumed a hydrostatically equilibrated shape of a rotating triaxial ellipsoid. In case of Enceladus, there is some ambiguity inherent in the data, and it cannot easily be distinguished between a homogeneous shape and, for instance, a 'frozen in' differentiated structure corresponding to a greater rotation rate in the past, when Enceladus was rotating synchronously closer to Saturn (Porco et al., 2006). Shape data are more consistent with a homogeneous structure (Thomas et al., 2007), which is on the other hand unexpected because of the activity and thermal state of the satellite. Differentiated structures would probably imply large, low-density cores. In addition, the layers might have lateral inhomogeneities in density. As noted by Thomas et al. (2007), fully relaxed, radially differentiated, laterally homogeneous models do not explain Enceladus' shape. A homogeneous structure would rule out a global ocean on Enceladus, which certainly would have implied complete differentiation. McKinnon (2013) suggested an irregular core, in terms of either density or topography, that would act to enhance the global geoid. Relaxation to the geoid could explain Enceladus' nonhydrostatic shape. Close flybys are required to further constrain Enceladus' internal structure. It should be noted that local processes, for example, a reorientation of an upwelling plume toward the south pole (Nimmo and Pappalardo, 2006), may contribute to the J_2 value. Thus, further flybys, equatorial and over the poles, would be required to reliably infer the structure of Enceladus.

The plumes observed by *Cassini* consist of a gas phase including mainly H_2O plus 5% CO_2, 1% CH_4, 1% NH_3, and small quantities of heavier hydrocarbons and organic molecules (Hansen et al., 2011; Waite et al., 2009) and a dust phase including water ice plus 1% salt, mainly NaCl (Postberg et al., 2011). The high water content can be interpreted as evidence for liquid water located close to the surface (Porco et al., 2006; Postberg et al., 2011; Schmidt et al., 2008). The required heat source to melt the ice and to power the activity, however, still remains enigmatic (see, e.g., Spencer and Nimmo (2013) and references therein).

10.18.7.5 Rhea

With a radius of 764.3 km, Rhea is the second largest satellite in the Saturn system. The density has been determined to be 1233 kg m^{-3} (Thomas et al., 2007), implying only a small amount (about 30%) of rock and a large fraction of water ice. Rhea's surface is heavily cratered. Prior to the *Cassini* flyby in November 2005, it was unclear if Rhea is differentiated, and hence, various internal structures had to be taken into account (Castillo-Rogez, 2006). The analysis of the Doppler data suggests an almost homogeneous structure (Anderson and Schubert, 2007, 2010) or a weakly differentiated satellite interior (Iess et al., 2007). Although there still remains a discrepancy in the value of C_{22} derived by Anderson and Schubert (2010) and Iess et al. (2007), a differentiated structure can be ruled out. In the latter analysis, the C_{22}/J_2 ratio is consistent with hydrostatic equilibrium within the given error bounds (Iess et al., 2007). A small degree of differentiation may be due to a concentration of rocks toward the center. Based on limb profiles, Nimmo et al. (2010) had estimated the long-wavelength topography of Rhea. Uncompensated degree-3 topography may contaminate the degree-2 gravity field data, a question that could be resolved by additional gravity flybys. Shape measurements of Rhea are consistent with the satellite being in a hydrostatic state (Thomas, 2010).

The possibility of Rhea being fully differentiated can be ruled out on the basis of the obtained data. This would most likely also rule out the existence of a deep subsurface ocean that would have been conceivable if the satellite were differentiated and contained at least small amounts of ammonia (Hussmann et al., 2006). In the models based on the *Cassini* Doppler data, the ice-I/ice-II phase transition occurs in the central region of the satellite. The formation of an ice-II core would be associated with a global contraction of the satellite that would have left its marks at the satellite's surface. However, being dominated by craters, the surface of Rhea is lacking such evidence. Whether the formation of an ice-II core occurred early in the satellite's evolution or whether temperatures were too low for the formation of ice-II remains unclear. According to Barr and Canup (2008), Rhea's partially differentiated state would imply slow accretion scenarios (~2 My) or a delayed accretion (by ~2–7 My) to avoid melting of the ices. It was also suggested that a high abundance of ammonia ice in Rhea's interior may have prevented the formation of the ice-II phase in the central regions (Prentice, 2006). Relaxation of impact crater topography suggests that Rhea has experienced heat flows of a few tens of m Wm^{-2} early in its history (White et al., 2013). Compared to Iapetus that is similar with respect to surface gravity, Rhea must have had a different thermal evolution with at least short phases of enhanced heating. The latter is probably related to rapid despinning and tidal interaction with Saturn (White et al., 2013).

10.18.7.6 Iapetus

With respect to size and density, Iapetus is comparable to Rhea. However, both satellites differ significantly in their surface characteristics as a consequence of their individual thermal and geologic evolution. Iapetus has a radius of 735.6 km and, based on its mean density of only 1088 kg m^{-3}, its rock content is even smaller than that of Rhea. Iapetus' interior is thus dominated by ice and the formation of ice-II in its central region is likely (Schubert et al., 1986). Long-lived radioactive isotopes in its small rock component are insufficient to completely differentiate the satellite (Ellsworth and Schubert, 1983; Multhaup and Spohn, 2007).

Iapetus' shape is inconsistent with its present synchronous rotation rate of almost 80 days and is best represented by an oblate spheroid (Thomas, 2010; Thomas et al., 2007). The difference between the equatorial and the polar radius corresponds to an equilibrium shape with a 16 h period. This suggests that Iapetus has preserved this shape as its rotation rate was decreasing to reach the stable state of synchronous rotation. In such scenarios, Iapetus formed at a high spin rate and achieved isostatic equilibrium by strong interior heating caused by short-lived radioactive isotopes shortly after formation (Castillo-Rogez et al., 2009, 2011; Robuchon et al., 2010). According to an alternative scenario (Sandwell and Schubert,

2010), Iapetus' shape could be a consequence of substantial porosity within the outer ice layer and subsequent compaction due to heating of long-lived radioactive isotopes. This scenario would also provide an explanation for Iapetus' equatorial ridge, a surface feature on Iapetus that is unique in the solar system. It stretches around almost the entire satellite with height up to 13–20 km (Giese et al., 2008a; Porco et al., 2005). The ridge is heavily cratered implying that it was formed early in the satellite's evolution. In the scenario by Sandwell and Schubert (2010), the ridge is a consequence of strain concentration along the equator during the process of compaction. However, formation scenarios for the equatorial ridge are still under discussion. Whereas a tectonic or volcanic origin seems to be unlikely (e.g., Singer and McKinnon, 2011), exogenic formation scenarios are promising. Ip (2006) suggested collisional accretion of an equatorial ring system shortly after satellite formation. However, in this scenario, the early time of formation is problematic because the satellite needs to be cold and rigid enough to sustain the massive load of the ridge. To overcome this, Levison et al. (2011) and Dombard et al. (2012) have suggested a giant impact and the formation of a small subsatellite and debris around Iapetus. Whereas in the scenario of Levison et al. (2011) the subsatellite eventually escapes from Iapetus, it is tidally disrupted in the scenario of Dombard et al. (2012). In both cases, the debris would form the ridge during a phase when Iapetus was already cold enough to sustain the ridge.

In the scenario of Levison et al. (2011), Iapetus is faster despun by the additional interaction with the subsatellite. Whether there is any close connection between despinning and the formation of the equatorial ridge is unclear. The formation and the preservation of this ridge are still enigmatic. The observations suggest that the interior was dissipative enough to allow for despinning and that the lithosphere was thick enough to support the mountain chain on a geologic timescale.

10.18.7.7 Triton

From the *Voyager* 2 flyby in 1989, Triton's radius, mass, and density were determined to be 1353.4 km (Thomas, 2000), 214×10^{20} kg, and 2054 kg m^{-3}, respectively (Tyler et al., 1989). However, on the basis of gravity measurements and from the shape solution derived from the analysis of limb profiles, the moment of inertia cannot be inferred (Thomas, 2000; Tyler et al., 1989), thus constraints on internal structure are sparse. The high density suggests a large rock-mass fraction of about 72%, implying substantial radiogenic heat sources in the deep interior. After Europa with a rock-mass fraction exceeding 90%, Triton has the highest density of all the icy satellites.

Because of its retrograde orbit around Neptune, it was suggested by McCord (1966) and later by McKinnon (1984) and McKinnon and Leith (1995) that Triton was captured by Neptune. As proposed by Agnor and Hamilton (2006), Triton was captured as a binary object, with the companion being lost during the capture process. Such models are motivated by Triton's high inclination of about 157° and its retrograde orbit, which are unique features among the large satellites in the solar system. With an eccentricity of 0.0004, Triton's orbit is almost an exact circle. Correia (2009) has shown how the orbit evolves from a highly eccentric one to the current circular one while Triton is captured into the synchronous rotation state. Whatever the capture mechanism was like, it involves high eccentricities shortly after capture (Ćuk and Gladman, 2005). Tidal dissipation in the satellite or gas drag within the Neptunian subnebula (McKinnon and Leith, 1995) may have efficiently damped the satellite's eccentricity, suggesting that at least Triton's early history was characterized by intense heating. However, the young surface age of about 0.1–0.35 Ga and 6–50 My regionally estimated from crater counts (Schenk and Zahnle, 2007; Stern and McKinnon, 2000; Zahnle et al., 2003) suggests that Triton was geologically active, even in its recent past. This is consistent with the diversity of geologic surface units imaged by *Voyager* 2 (Smith et al., 1989). The detection of geyser-like plumes at the surface (Smith et al., 1989; Soderblom et al., 1990) further revealed that Triton is one of only three satellites in the solar system showing present-day eruptive activity, the other two being Io and Enceladus. However, in case of Triton, the observed activity is believed to be driven by solar insolation during seasons of permanent sunlight. However, geothermal heat may additionally contribute to the required energy budget (Brown et al., 1990).

The process of being captured by Neptune would lead to strong internal heating as the orbit evolved, most probably implying differentiation of the satellite. This would result in a structure consisting of a rock core and an H_2O layer. The physical state of the latter remains unclear due to the poorly constrained abundance of volatiles and temperature profile in the shell. A subsurface ocean would be conceivable if small amounts of ammonia are present in the system (McKinnon and Leith, 1995). Based on thermal equilibrium models, and considering a pure ice-I layer above a water–ammonia ocean, the ice layer would be about 140–200 km thick covering an ammonia–water ocean of about 150 km thickness (Hussmann et al., 2006). Using estimates of the heat flow, required for explaining tectonic features on Triton, Ruiz (2003) estimated an ice-I layer thickness of only 20 km. Due to the lack of data on gravity and composition, it is also unclear if the ocean, if it exists, would be enclosed between ice-I and high-pressure ice. Thus, a 'Europa-like' structure consisting of a large silicate core, a water–ammonia ocean, and an ice-I shell cannot be ruled out. Such a structure would be intriguing because of its possible astrobiological consequences. Gaeman et al. (2012) have investigated the stability of an ocean during the phase of orbital eccentricity damping. Depending on the initial values of eccentricity, an ocean may be sustained over 4.5 Gy. If circularization of Triton's orbit is slow enough, a thin NH_3-rich water ocean might exist even today.

Whatever the past thermal history, it has permitted outgassing of enough nitrogen and methane to produce surface ice deposits that are moved seasonally from one pole to the other and maintain a thin microbar atmosphere of mostly nitrogen (Lunine and Nolan, 1992).

10.18.8 Future Prospects for Determining Satellite Internal Structure

At the time of writing of this chapter, the *Cassini* orbiter has completed 9 years of operation in Saturn orbit, more than twice the duration of its prime mission. A 2-year extended 'equinox' mission studied Titan, the other moons, rings, and

Saturn atmosphere under conditions of equinox solar illumination. A 6-year 'solstice' mission began in 2011 with the intention of operating the spacecraft until fuel exhaustion demands a destructive entry into Saturn's atmosphere in 2017. The solstice mission is allowing gravity mapping of Enceladus and more thorough gravity mapping of Titan.

Juno, a polar orbiting spacecraft launched to Jupiter in August 2011, will provide detailed information on the internal structure and perhaps the dissipation factor Q_l of Jupiter, which will help in understanding tidal interactions between the large moons, mainly Io, and Jupiter. *New Horizons* will fly by Pluto and its moon Charon in 2015, providing, however, limited information on internal structure.

ESA is preparing JUICE, the JUpiter ICy moons Explorer (Grasset et al., 2013), to orbit Ganymede and investigate Europa and Callisto by flybys. The mission will provide details on Ganymede's interior structure including higher moments of the gravity field and tidal variations that would be indicative for ocean detection and its characterization. The mission will also carry a subsurface radar instrument to investigate the upper layers (~9 km) of the ice shells of Ganymede, Europa, and Callisto. Further plans to measure in detail Europa's internal structure and thickness of its ice crust (Wahr et al., 2006) remain in limbo. JUICE will not arrive at Jupiter any earlier than 2030. Further progress on the internal structure of icy moons can be made in laboratory studies of the relevant materials, especially with respect to the physical properties of water ice mixed with other ices. High-pressure ices and clathrates could be investigated for various compositions. Thermophysical properties of interest for the evolution of icy satellites include the density, thermal expansivity, heat capacity, and thermal conductivity as a function of temperature, pressure, and composition. For tidal deformation processes and convective heat transport, the rheological properties should be investigated on different frequencies. Main parameters, crucial for the evolution of icy satellites, include rigidity, viscosity, and transient creep parameters as a function of temperature and frequency of the forcing.

Acknowledgments

The authors would like to thank Michael Bland and an anonymous reviewer for the very detailed and constructive comments on the manuscript. JIL acknowledges support from the *Cassini* Project in preparing the manuscript. Part of this work was supported by NASA's Astrobiology Institute node 'Icy Worlds' and was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

Agnor CB and Hamilton DP (2006) Neptune's capture of its moon Triton in a binary-planet gravitational encounter. *Nature* 441: 192–194.

Anderson JD, Jacobson RA, Lau EL, Moore WB, and Schubert G (2001a) Io's gravity field, and interior structure. *Journal of Geophysical Research* 106: 32963–32970.

Anderson JD, Jacobson RA, McElrath TP, Moore WB, Schubert G, and Thomas PC (2001b) Shape, mean radius, gravity field, and interior structure of Callisto. *Icarus* 153: 157–161.

Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996) Gravitational constraints on the internal structure of Ganymede. *Nature* 384: 541–543.

Anderson JD and Schubert G (2007) Saturn's satellite Rhea is a homogeneous mix of rock and ice. *Geophysical Research Letters* 34: L02202. http://dx.doi.org/10.1029/2006GL028100.

Anderson JD and Schubert G (2010) Rhea's gravitational field and interior structure inferred from archival data files of the 2005 Cassini flyby. *Physics of the Earth and Planetary Interiors* 178: 176–182.

Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998) Europa's differentiated structure, inferences from four Galileo encounters. *Science* 281: 2019–2022.

Asplund M, Grevesse N, Sauval AJ, and Scott P (2009) The chemical composition of the Sun. *Annual Review of Astronomy and Astrophysics* 47: 481–522.

Atreya SK, Adams EY, Niemann HB, et al. (2006) Titan's methane cycle. *Planetary and Space Science* 54: 1177–1187.

Barr AC (2008) Mobile lid convection beneath Enceladus' south polar terrain. *Journal of Geophysical Research* 113(E7): E07009.

Barr AC and Canup RM (2008) Constraints on gas giant satellite formation from the interior states of partially differentiated satellites. *Icarus* 198: 163–177.

Barr AC and Canup RM (2010) Origin of the Ganymede–Callisto dichotomy by impacts during the late heavy bombardment. *Nature Geoscience* 3: 164–167.

Barr AC, Citron RI, and Canup RM (2010) Origin of a partially differentiated Titan. *Icarus* 209: 858–862.

Barr AC and McKinnon WB (2007) Convection in ice I shells and mantles with self-consistent grain size. *Journal of Geophysical Research* 112(E2): E02012.

Barr AC, Pappalardo RT, and Zhong S (2004) Convective instability in ice-I with non-Newtonian rheology: Application to the ice Galilean satellites. *Journal of Geophysical Research* 109: E12008. http://dx.doi.org/10.1029/2004JE002296.

Barr AC and Showman AP (2009) Heat transfer in Europa's icy shell. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 405–430. Tucson, AZ: University of Arizona Press.

Beghin C, Randriamboarison O, Hamelin M, et al. (2012) Analytic theory of Titan's Schumann resonance: Constraints on ionospheric conductivity and buried water ocean. *Icarus* 218: 1028–1042.

Behounkova M, Tobie G, Choblet G, and Cadek O (2012) Tidally-induced melting events as the origin of south-pole activity on Enceladus. *Icarus* 219: 655–664.

Bierhaus EB, Zahnle K, and Chapman CR (2009) Europa's crater distributions and surface ages. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 353–368. Tucson, AZ: University of Arizona Press.

Bills BG and Nimmo F (2008) Forced obliquity and moments of inertia of Titan. *Icarus* 196: 293–297.

Bills BG and Nimmo F (2011) Rotational dynamics and internal structure of Titan. *Icarus* 214: 351–355.

Bland MT, Showman AP, and Tobie G (2008) The production of Ganymede's magnetic field. *Icarus* 198: 384–399.

Bland MT, Showman AP, and Tobie G (2009) The orbital–thermal evolution and global expansion of Ganymede. *Icarus* 200: 207–221.

Bland MT, Singer KN, McKinnon WB, and Schenk PM (2012) Enceladus' extreme heat flux as revealed by its relaxed craters. *Geophysical Research Letters* 39(17): L17204.

Bockelée-Morvan D, Crovisier J, Mumma MJ, and Weaver HA (2004) The composition of cometary volatiles. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 391–423. Tucson, AZ: University of Arizona Press.

Brown RH, Clark RN, Buratti BJ, et al. (2006) Composition and physical properties of Enceladus' surface. *Science* 311: 1425–1428.

Brown RH, Cruikshank DP, Veverka J, Helfenstein P, and Eluszkiewicz J (1995) Surface composition and photometric properties of Triton. In: Cruikshank DP (ed.) *Neptune and Triton*, pp. 991–1030. Tucson, AZ: University of Arizona Press.

Brown RH, Johnson TV, Kirk RL, and Soderblom LA (1990) Energy sources for Triton's geyser-like plumes. *Science* 250: 431–435.

Calvin WM and Clark RN (1991) Modeling the reflectance spectrum of Callisto 0.25 to 4.1 μm. *Icarus* 89: 305–317.

Canup RM and Ward WR (2002) Formation of the Galilean satellites: Conditions of accretion. *The Astronomical Journal* 124: 3404–3423.

Canup RM and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839.

Carr MH, Belton MJS, Chapman CR, et al. (1998) Evidence for a subsurface ocean on Europa. *Nature* 391: 363–365.

Cartwright R, Emery J, Rivkin A, and Trilling D (2013) Near-infrared spectroscopy of Uranian satellites: Searching for carbon dioxide on Umbriel, Titania, and Oberon. *Lunar and Planetary Science Conference* 44: Abstract 1195.

Cassen P, Peale SJ, and Reynolds RT (1980) Tidal dissipation in Europa – A correction. *Geophysical Research Letters* 7: 987–988.

Cassen P, Reynolds RT, and Peale SJ (1979) Is there liquid water on Europa? *Geophysical Research Letters* 6: 731–734.

Castillo-Rogez J (2006) Internal structure of Rhea. *Journal of Geophysical Research* 111: E11005. http://dx.doi.org/10.1029/2004JE002379.

Castillo-Rogez JC, Efroimsky M, and Lainey V (2011) The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. *Journal of Geophysical Research* 116(E9): E09008.
Castillo-Rogez J, Johnson TV, Lee M-H, Turner NJ, Matson DL, and Lunine JI (2009) ^{26}Al decay: Heat production and a revised age for Iapetus. *Icarus* 204: 658–662.
Castillo-Rogez J and Lunine JI (2010) Evolution of Titan's rocky core constrained by Cassini observations. *Geophysical Research Letters* 37: L20205.
Castillo-Rogez JC, Matson DL, Sotin C, Johnson TV, Lunine JI, and Thomas PC (2007) Iapetus' geophysics: Rotation rate, shape, and equatorial ridge. *Icarus* 190: 179–202.
Chen EMA, Nimmo F, and Glatzmaier GA (2013) Tidal heating in icy satellite oceans. *Icarus* 229: 11–30.
Chizhov VE (1993) Thermodynamic properties and thermal equation of state of high-pressure ice phases. *Prikladnaya Mekhanika i Tekhnicheskaya Fizika* 2: 113–123.
Choukroun M and Grasset O (2010) Thermodynamic data and modeling of the water and water–ammonia phase diagrams for planetary geophysics. *Journal of Chemical Physics* 133: 144502.
Clark RN (1980) Ganymede, Europa, Callisto, and Saturn's rings: Compositional analysis from reflectance spectroscopy. *Icarus* 44: 388–409.
Clark RN, Brown RH, Jaumann R, et al. (2005) The surface composition of Phoebe as seen by the Cassini visual and infrared mapping spectrometer. *Nature* 435: 66–69.
Clark RN, Curchin JM, Barnes JW, et al. (2010) Detection and mapping of hydrocarbon deposits on Titan. *Journal of Geophysical Research* 115: E10005, 28 pp.
Collins GC and Goodman JC (2007) Enceladus' south polar sea. *Icarus* 189: 72–82.
Collins GC, Head JW, Pappalardo RT, and Spaun NA (2000) Evaluation of models for the formation of chaotic terrain on Europa. *Journal of Geophysical Research* 105: 1709–1716.
Collins G and Nimmo F (2009) Chaotic terrain on Europa. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 259–282. Tucson, AZ: University of Arizona Press.
Consolmagno GJ and Lewis JS (1976) Structural and thermal models of icy Galilean Satellites. In: Gehrels TA (ed.) *Jupiter*, pp. 1035–1051. Tucson, AZ: University of Arizona Press.
Consolmagno GJ and Lewis JS (1978) The evolution of icy satellite interiors and surfaces. *Icarus* 34: 280–293.
Coradini A, Federico C, Forni O, and Magni G (1995) Origin and thermal evolution of icy satellites. *Surveys in Geophysics* 16: 533–591.
Correia ACM (2009) Secular evolution of a satellite by tidal effect: Application to Triton. *The Astronomical Journal* 704: L1–L4.
Coustenis A, Bézard B, and Gautier D (1989) Titan's atmosphere from Voyager infrared observations I. The gas composition of Titan's equatorial region. *Icarus* 80: 54–76.
Crary FJ and Bagenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research: Planets* 103: 25757–25773.
Ćuk M and Gladman BJ (2005) Constraints on the orbital evolution of Triton. *The Astronomical Journal* 626: L113–L116.
Cyr KE, Sharp CM, and Lunine JI (1999) Effects of the redistribution of water in the solar nebula on nebular chemistry. *Journal of Geophysical Research* 104(E8): 19003–19014.
Dartois E and d'Hendecourt L (2001) Search for ammonia ice in cold dust envelopes around YSO's. *Astronomy and Astrophysics* 365: 144–165.
Davaille A and Jaupart C (1993) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
Dermott SF (1979) Shapes and gravitational moments of satellites and asteroids. *Icarus* 37: 575–586.
Deschamps F and Sotin C (2001) Thermal convection in the outer shell of large icy satellites. *Journal of Geophysical Research* 106: 5107–5121.
Dodson-Robinson S, Willacy K, Bodenheimer P, Turner NJ, and Beichman CA (2009) Ice lines, planetesimal composition and solid surface density in the solar nebula. *Icarus* 200: 672–693.
Doggett T, Greeley R, Figueredo P, and Tanaka K (2009) Geologic stratigraphy and evolution of Europa's surface. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 259–282. Tucson, AZ: University of Arizona Press.
Dombard AJ, Cheng AF, McKinnon WB, and Kay JP (2012) Delayed formation of the equatorial ridge on Iapetus from a subsatellite created in a giant impact. *Journal of Geophysical Research* 117(E3): E03002.
Ducci M, Iess L, Armstrong JW, et al. (2012) The geodesy of the main saturnian satellites from range rate measurements of the Cassini spacecraft. In: *43rd Lunar and Planetary Science Conference*, Texas: The Woodlands, Abstract 2200.
Dumoulin C, Doin M-P, and Fleitout L (1999) Heat transport in stagnant lid convection with temperature- and pressure-dependent Newtonian and non-Newtonian rheology. *Journal of Geophysical Research* 104: 12759–12777.
Durham WB, McKinnon WB, and Stern LA (2005) Cold compaction of water ice. *Geophysical Research Letters* 32(18): L18202.
Durham WB, Prieto-Ballesteros O, Goldsby DL, and Kargel JS (2010) Rheological and thermal properties of icy materials. *Space Science Reviews* 153: 273–298.
Durham WB and Stern LA (2001) Rheological properties of water ice – Applications to satellites of the outer planets. *Annual Review of Earth and Planetary Sciences* 29: 295–330.
Elachi C, Wall S, Allison M, et al. (2005) First views of the surface of Titan from the Cassini RADAR. *Science* 308: 970–974.
Ellsworth K and Schubert G (1983) Saturn's icy satellites – Thermal and structural models. *Icarus* 54: 490–510.
Estrada PR and Mosqueira I (2006) A gas-poor planetesimal capture model for the formation of giant planet satellite systems. *Icarus* 181: 486–509.
Farinella P, Mazari F, and Matteoli S (1997) The disruption of Hyperion and the origin of Titan's atmosphere. *The Astronomical Journal* 113: 2312–2316.
Figueredo PH and Greeley R (2004) Resurfacing history of Europa from pole-to-pole geological mapping. *Icarus* 167: 287–312.
Filacchione G, Capaccioni F, Clark RN, et al. (2013) The radial distribution of water ice and chromophores across Saturn's system. *Astrophysical Journal* 766: 76.
Fischer H-J and Spohn T (1990) Thermal–orbital histories of viscoelastic models of Io (J1). *Icarus* 83: 39–65.
Flasar F, Achterberg RK, Conrath BJ, et al. (2005) Titan's atmospheric temperatures, winds, and composition. *Science* 308: 975–978.
Fortes AD (2012) Titan's internal structure and evolutionary consequences. *Planetary and Space Science* 60: 10–17.
Gaeman J, Hier-Majumder S, and Roberts JH (2012) Sustainability of a subsurface ocean within Triton's interior. *Icarus* 220: 339–347.
Gao P and Stevenson DJ (2012) How does nonhydrostaticity affect the determination of the moment of inertia? *Lunar and Planetary Science Conference* 43: Abstract 1701.
Geissler PE, Greenberg R, Hoppa G, et al. (1998) Evidence for non-synchronous rotation on Europa. *Nature* 391: 368–370.
Giese B, Denk T, Neukum G, et al. (2008a) The topography of Iapetus' leading side. *Icarus* 193: 359–371.
Giese B, Wagner R, Hussmann H, et al. (2008b) Enceladus: An estimate of heat flux and lithospheric thickness from flexurally supported topography. *Geophysical Research Letters* 35: L240204.
Goldsby DL and Kohlstedt DL (2001) Superplastic deformation of ice: Experimental observations. *Journal of Geophysical Research* 106: 11017–11030.
Goodman JC and Lenferink E (2012) Numerical simulations of marine hydrothermal plumes for Europa and other icy worlds. *Icarus* 221: 970–983.
Grasset O and Pargamin J (2005) The ammonia–water system at high pressures: Implications for the methane of Titan. *Planetary and Space Science* 53: 371–384.
Grasset O and Parmentier EM (1998) Thermal convection in a volumetrically heated, infinite Prandtl number fluid with strongly temperature-dependent viscosity: Implications for planetary evolution. *Journal of Geophysical Research* 103: 18171–18181.
Grasset O and Sotin C (1996) The cooling rate of a liquid shell in Titan's interior. *Icarus* 123: 101–112.
Grasset O, Sotin C, and Deschamps F (2000) On the internal structure and dynamics of Titan. *Planetary and Space Science* 48: 617–636.
Grasset O, Dougherty MK, Coustenis A, et al. (2013) Jupiter Icy moons Explorer (JUICE): An ESA mission to orbit Ganymede and to characterize the Jupiter system. *Planetary and Space Science* 78: 1–21.
Greeley R, Chyba CF, Head JW III, et al. (2004) In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter. The Planet, Satellites, and Magnetosphere*, pp. 329–336. Cambridge: Cambridge University Press.
Greeley R, Sullivan R, Klemaszewski J, et al. (1998) Europa: Initial Galileo geological observations. *Icarus* 135: 4–24.
Greenberg R (1982) Orbital evolution of the Galilean Satellites. In: Morrison D (ed.) *Satellites of Jupiter*, pp. 65–92. Tucson, AZ: University of Arizona Press.
Greenberg R, Croft SK, Janes DM, et al. (1991) Miranda. In: Bergstralh JT, Miner ED, and Matthews MS (eds.) *Uranus*, pp. 693–735. Tucson, AZ: University of Arizona Press.
Greenberg R, Geissler P, Hoppa G, et al. (1998) Tectonic Processes on Europa: Tidal stresses, mechanical response, and visible features. *Icarus* 135: 64–78.
Greenberg R, Geissler P, Tufts BR, and Hoppa GV (2000) Habitability of Europa's crust: The role of tidal-tectonic processes. *Journal of Geophysical Research* 105(E7): 17,551–17,562.
Greenberg R, Hoppa GV, Tufts BR, Geissler P, Riley J, and Kadel S (1999) Chaos on Europa. *Icarus* 141: 263–286.
Grott M, Sohl F, and Hussmann H (2007) Degree-one convection and the origin of Enceladus' dichotomy. *Icarus* 191: 203–210.
Han L, Tobie G, and Showman AP (2012) The impact of a weak south pole on thermal convection in Enceladus' ice shell. *Icarus* 218: 320–330.

Hand K and Brown ME (2013) Keck II observations of hemispherical differences in H_2O_2 on Europa. *The Astrophysical Journal Letters* 766: L21.

Hand K and Chyba CF (2007) Empirical constraints on the salinity of the European ocean and implications for a thin ice shell. *Icarus* 189: 424–438.

Hand KP, Chyba CF, Priscu JC, Carlson RW, and Nealson KH (2009) Astrobiology and the potential for life on Europa. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 589–630. Tucson, AZ: University of Arizona Press.

Hansen CJ, Shemansky DE, Esposito LW, et al. (2011) The composition and structure of the Enceladus plume. *Geophysical Research Letters* 38(11): L11202.

Hedman MM, Gosmeyer CM, Nicholson PD, et al. (2013) An observed correlation between plume activity and tidal stresses on Enceladus. *Nature* 500: 182–184.

Hibbits CA, Pappalardo RT, Hansen GB, and McCord TB (2003) Carbon dioxide on Ganymede. *Journal of Geophysical Research* 108(E5): 5036. http://dx.doi.org/10.1029/2002JE001956.

Hobbs PV (1974) *Ice Physics*. Oxford: Clarendon Press.

Hogenboom DL, Kargel JS, Consolmagno GJ, Holden TC, Lee L, and Buyyounouski M (1997) The ammonia–water system and the chemical differentiation of icy satellites. *Icarus* 128: 171–180.

Hoppa GV, Tufts BR, Greenberg R, and Geissler PE (1999) Formation of cycloidal features on Europa. *Science* 285: 1899–1902.

Howett CJA, Spencer JR, Pearl J, and Segura M (2011) High heat flow from Enceladus' south polar region measured using 10–600 cm^{-1} Cassini/CIRS data. *Journal of Geophysical Research* 116(E3): E03003.

Hubbard WB and Anderson JD (1978) Possible flyby measurements of Galilean satellite interior structure. *Icarus* 33: 336–341.

Hunten DM, Tomasko MG, Flasar FM, Samuelson RE, Strobel DF, and Stevenson DJ (1984) Titan. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 671–759. Tucson, AZ: University of Arizona Press.

Hurford TA, Bills BG, Helfenstein P, Greenberg R, Hoppa GV, and Hamilton DP (2009) Geological implications of a physical libration on Enceladus. *Icarus* 203: 541–552.

Hurford TA, Helfenstein P, Hoppa GV, Greenberg R, and Bills BG (2007) Eruptions arising from tidally controlled periodic openings of rifts on Enceladus. *Nature* 447: 292–294.

Hussmann H, Choblet G, Lainey V, et al. (2010) Implications of rotation, orbital states, energy sources, and heat transport for internal processes in icy satellites. *Space Science Reviews* 153: 317–348.

Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.

Hussmann H and Spohn T (2004) Thermal–orbital evolution of Io and Europa. *Icarus* 171: 391–410.

Hussmann H, Spohn T, and Wiezcerkowski K (2002) Thermal equilibrium states of Europa's ice shell: Implications for internal ocean thickness and surface heat flow. *Icarus* 156: 143–151.

Iess L, Jacobson RA, Ducci M, et al. (2012) The tides of Titan. *Science* 337: 457–459.

Iess L, Rappaport NJ, Jacobson RA, et al. (2010) Gravity field, shape and moment of inertia of Titan. *Science* 327: 1367–1369.

Iess LNJ, Rappaport P, Tortora J, et al. (2007) Gravity field and interior of Rhea from Cassini data analysis. *Icarus* 190: 585–593.

Ip W-H (2006) On a ring origin of the equatorial ridge of Iapetus. *Geophysical Research Letters* 33: L16203.

Jacobson RA (2004) The orbits of the major Saturnian satellites and the gravity field of Saturn from spacecraft and Earth-based observations. *The Astronomical Journal* 128: 492–501.

Jacobson RA, Antreasian PG, Bordi JJ, et al. (2006) The gravity field of the Saturnian system from satellite observations and spacecraft tracking data. *The Astronomical Journal* 132: 2520–2526.

Johnson TV and Estrada PR (2009) Origin of the Saturn system. In: Dougherty M, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 55–74. Dordrecht: Springer.

Johnson TV and Lunine JI (2005) Saturn's moon Phoebe as a captured body from the outer solar system. *Nature* 435: 69–71.

Jara-Orue H and Vermeersen LLA (2014) The forced libration of Europa's deformable shell and its dependence on interior parameters. *Icarus* 229: 31–44.

Kargel JS, Kaye JZ, Head JW, et al. (2000) Europa's crust and ocean: Origin, composition, and the prospects for life. *Icarus* 148: 226–265.

Kattenhorn SA and Hurford T (2009) Tectonics of Europa. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 199–236. Tucson, AZ: University of Arizona Press.

Kaula WM (1968) *An Introduction to Planetary Physics: The Terrestrial Planets, Space Science Text Series*. New York: Wiley.

Kempf S, Beckmann U, and Schmidt J (2010) How the Enceladus dust plume feeds Saturn's E-ring. *Icarus* 206: 446–457.

Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.

Kimura J, Nakagawa T, and Kurita K (2009) Size and compositional constraints of Ganymede's metallic core for driving an active dynamo. *Icarus* 202: 216–224.

Kirk RL and Stevenson DJ (1987) Thermal evolution of a differentiated Ganymede and implications for surface features. *Icarus* 69: 91–134.

Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.

Kivelson MG, Khurana KK, Krishan K, et al. (2000) Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.

Kivelson MG, Warnecke J, Bennett L, et al. (1998) Ganymede's magnetosphere: Magnetometer overview. *Journal of Geophysical Research: Planets* 103: 19963–19972.

Klinger J (1981) Some consequences of a phase transition of water ice on the heat balance of comet nuclei. *Icarus* 47: 320–324.

Kuramoto K and Matsui T (1994) Formation of a hot proto-atmosphere on the accreting giant icy satellite: Implications for the origin and evolution of Titan, Ganymede, and Callisto. *Journal of Geophysical Research* 99: 21183–21200.

Kuskov OL and Kronrod VA (2005) Internal structure of Europa and Callisto. *Icarus* 177: 550–569.

Lebreton J-P, Witasse O, Sollazzo C, et al. (2005) An overview of the descent and landing of the Huygens probe on Titan. *Nature* 438: 758–764.

Leliwa-Kopystyński J, Maruyama M, and Nakajima T (2002) The water–ammonia phase diagram up to 300 MPa: Application to icy satellites. *Icarus* 159: 518–528.

Levison HF, Walsh KJ, Barr AC, and Dones L (2011) Ridge formation and de-spinning of Iapetus via an impact-generated satellite. *Icarus* 214: 773–778.

Lewis JS (1971) Satellites of the outer planets: Their physical and chemical nature. *Icarus* 15: 175–185.

Lewis JS (1972) Low temperature condensation from the solar nebula. *Icarus* 16: 241–252.

Lindal GF, Wood GE, Hotz HB, Sweetnam DN, Eshleman VR, and Tyler GL (1983) The atmosphere of Titan – An analysis of the Voyager 1 radio occultation measurements. *Icarus* 53: 348–363.

Lissauer JJ, Peale SJ, and Cuzzi JN (1984) Ring torque on Janus and the melting of Enceladus. *Icarus* 58: 159–168.

Lopes RMC, Stofan ER, Peckyno R, et al. (2010) Distribution and interplay of geologic processes on Titan from Cassini Radar data. *Icarus* 205: 540–558.

Lorenz R and Lunine J (2005) Titan's surface before Cassini. *Planetary and Space Science* 53: 557–576.

Lowell RP and DuBose M (2005) Hydrothermal systems on Europa. *Geophysical Research Letters* 32: L05202.

Lunine JI and Nolan MC (1992) A massive early atmosphere on Triton. *Icarus* 100: 221–234.

Lunine JI and Stevenson DJ (1982) Formation of the Galilean satellites in a gaseous nebula. *Icarus* 52: 14–39.

Lunine JI and Stevenson DJ (1987) Clathrate and ammonia hydrates at high pressure – Application to the origin of methane on Titan. *Icarus* 70: 61–77.

Lupo MJ (1982) Mass-radius relationships in icy satellites after Voyager. *Icarus* 52: 40–53.

Matson DL, Castillo-Rogez JC, Schubert G, Sotin C, and McKinnon WB (2009) The thermal evolution and internal structure of Saturn's mid-sized icy satellites. In: Dougherty M, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 577–612. Dordrecht: Springer.

McCord TB (1966) Dynamical evolution of the Neptunian system. *The Astronomical Journal* 71: 585.

McKinnon WB (1984) On the origin of Triton and Pluto. *Nature* 311: 355–358.

McKinnon WB (1999) Convective instability in Europa's floating ice shell. *Geophysical Research Letters* 26: 951–954.

McKinnon WB (2006) On convection in ice I shells of outer solar system bodies, with detailed application to Callisto. *Icarus* 183: 435–450.

McKinnon WB (2013) The shape of Enceladus as explained by an irregular core: Implications for gravity, libration, and survival of its subsurface ocean. *Journal of Geophysical Research* 118: 1–14.

McKinnon WB and Leith AC (1995) Gas drag and the orbital evolution of a captured Triton. *Icarus* 118: 392–413.

Meriggiola R and Iess L (2012) *A new rotational model of Titan from Cassini SAR data*, EPSC 2012, Abstract id. EPSC2012–593.

Merk R, Breuer D, and Spohn T (2002) Numerical modelling of ^{26}Al-induced radioactive melting of asteroids considering accretion. *Icarus* 159: 183–191.

Meyer J and Wisdom J (2007a) Tidal heating in Enceladus. *Icarus* 188: 535–539.

Meyer J and Wisdom J (2007b) Tidal evolution of Mimas, Enceladus, and Dione. *Icarus* 193: 213–223.
Meyer J and Wisdom J (2008) Episodic volcanism on Enceladus. *Icarus* 188: 535–539.
Mitri G and Showman AP (2005) Convective-conductive transitions and sensitivity of a convecting ice shell to perturbations in heat flux and tidal heating rate: Implications for Europa. *Icarus* 177: 447–460.
Moore WB and Hussmann H (2009) Thermal evolution of Europa's silicate shell. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 369–380. Tucson, AZ: University of Arizona Press.
Moore WB and Schubert G (2000) The tidal response of Europa. *Icarus* 147: 317–319.
Moore WB and Schubert G (2003) The tidal response of Ganymede and Callisto with and without liquid water oceans. *Icarus* 166: 223–226.
Mosqueira I, Estrada P, and Turrini D (2010) Planetesimals and satellitesimals: Formation of the satellite systems. *Space Science Reviews* 153: 431–446.
Mousis O, Lunine JI, Thomas C, et al. (2009a) Clathration of volatiles in the solar nebula and implications for the origin of Titan's atmosphere. *Astrophysical Journal* 691: 1780–1786.
Mousis O, Lunine JI, Waite JH Jr., et al. (2009b) Formation conditions of Enceladus and origin of its methane reservoir. *The Astrophysical Journal Letters* 701: L39–L42.
Mueller S and McKinnon WB (1988) Three-layered models of Ganymede and Callisto: Compositions, structures and aspects of evolution. *Icarus* 76: 437–464.
Multhaup K and Spohn T (2007) Stagnant lid convection in the mid-sized icy satellites of Saturn. *Icarus* 186: 420–435.
Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
Neubauer FM (1999) Alfvén wings and electromagnetic induction in the interiors: Europa and Callisto. *Journal of Geophysical Research* 104(A12): 28671.
Niemann HB, Atreya SK, Bauer SJ, et al. (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438: 779–784.
Nimmo F, Bills BG, Thomas PC, and Asmar SW (2010) Geophysical implications of the long-wavelength topography of Rhea. *Journal of Geophysical Research* 115(E10): E10008.
Nimmo F and Manga M (2002) Causes, characteristics and consequences of convective diaparism on Europa. *Geophysical Research Letters* 29: 24.
Nimmo F and Manga M (2009) Geodynamics of Europa's Icy Shell. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 381–404. Tucson, AZ: University of Arizona Press.
Nimmo F and Pappalardo RT (2006) Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* 441: 614–616.
Nimmo F, Spencer JR, Pappalardo RT, and Mullen ME (2007) Shear heating as the origin of the plumes and heat flux of Enceladus. *Nature* 447: 289–291.
O'Neill C and Nimmo F (2010) The role of episodic overturn in generating the surface geology and heat flow on Enceladus. *Nature Geoscience* 3: 88–91.
O'Rourke JG and Stevenson DJ (2013) *Stability of ice/rock mixtures with application to a partially differentiated Titan*, arXiv:1210.5280v1.
Ojakangas GW and Stevenson DJ (1986) Episodic volcanism of tidally heated satellites with application to Io. *Icarus* 66: 341–358.
Ojakangas GW and Stevenson DJ (1989) Thermal state of an ice shell on Europa. *Icarus* 81: 220–241.
Olkin CB, Wasserman LH, and Franz OG (2003) The mass ratio of Charon to Pluto from Hubble Space Telescope astrometry with the fine guidance sensors. *Icarus* 164: 254–259.
Owen T (1982) The composition and origin of Titan's atmosphere. *Planetary and Space Science* 30: 833–838.
Owen T (2000) On the origin of Titan's atmosphere. *Planetary and Space Science* 48: 747–752.
Palguta J, Schubert G, Zhang K, and Anderson JD (2009) Constraints on the location, magnitude, and dimensions of Ganymede's mass anomalies. *Icarus* 201: 615–625.
Pappalardo RT and Barr AC (2004) The origin of domes on Europa: The role of thermally induced compositional diapirism. *Geophysical Research Letters* 31(1): L01701.
Pappalardo RT, Head JW, Greeley R, et al. (1998) Geologic evidence for solid-state convection in Europa's ice shell. *Nature* 391: 365–368.
Pappalardo RT, Belton MJS, Brenemann HH, et al. (1999) Does Europa have a subsurface ocean? Evaluation of the geological evidence. *Journal of Geophysical Research* 104: 24015–24055.
Patterson GW, Collins GC, Head JW, et al. (2010) Global geological mapping of Ganymede. *Icarus* 207: 845–867.
Patthoff DA and Kattenhorn SA (2011) A fracture history on Enceladus provides evidence for a global ocean. *Geophysical Research Letters* 38: L18201.
Peale SJ (1986) Orbital resonances, unusual configurations and exotic rotation states among planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 159–223. Tucson, AZ: University of Arizona Press.
Peale SJ (2003) Tidally induced volcanism. *Celestial Mechanics and Dynamical Astronomy* 87: 129–155.
Peale SJ, Cassen P, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.
Peale SJ and Lee MH (2002) A primordial origin of the Laplace relation among the Galilean satellites. *Science* 298: 593–597.
Poirier JP, Boloh L, and Chambon P (1983) Tidal dissipation in small viscoelastic ice moons – The case of Enceladus. *Icarus* 55: 218–230.
Porco CC, Baker E, Barbara J, et al. (2005) Cassini imaging science: Initial results on Phoebe and Iapetus. *Science* 307: 1237–1242.
Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.
Postberg F, Schmidt J, Hillier J, Kempf S, and Srama R (2011) A salt-water reservoir as the source of a compositionally-stratified plume on Enceladus. *Nature* 474: 620–622.
Prentice AJR (2006) Saturn's icy moon Rhea: A prediction for its bulk chemical composition and physical structure at the time of the Cassini spacecraft first flyby. *Publications of the Astronomical Society of Australia* 23: 1–11.
Prockter LM and Patterson GW (2009) Morphology and evolution of Europa's ridges and bands. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 237–258. Tucson, AZ: University of Arizona Press.
Rappaport NJ, Giampieri G, and Anderson JD (2001) Perturbations of a spacecraft orbit during a hyperbolic flyby. *Icarus* 150: 168–180.
Rappaport NJ, Iess L, Tortora P, et al. (2007) Mass and interior of Enceladus from Cassini data analysis. *Icarus* 190: 175–178.
Rathbun JA, Musser GS, and Squyres SW (1998) Ice diapers on Europa: Implications for liquid water. *Geophysical Research Letters* 25: 4157–4160.
Reynolds RT, Squyres SW, Colburn DS, and McKay CP (1983) On the habitability of Europa. *Icarus* 56: 246–254.
Riley J, Hoppa GV, Greenberg R, Tufts BR, and Geissler P (2000) Distribution of chaotic terrain on Europa. *Journal of Geophysical Research* 105: 22599–22615.
Roberts JH and Nimmo F (2008) Tidal heating and the long-term stability of a subsurface ocean on Enceladus. *Icarus* 194: 675–689.
Robuchon G, Choblet G, Tobie G, Cadek O, Sotin C, and Grasset O (2010) Coupling of thermal evolution and despinning of early Iapetus. *Icarus* 207: 959–971.
Ross MN and Schubert G (1987) Tidal heating in an internal ocean model of Europa. *Nature* 325: 132–134.
Ross MN and Schubert G (1989) Viscoelastic models of tidal heating in Enceladus. *Icarus* 78: 90–101.
Ruiz J (2001) The stability against freezing of an internal liquid-water ocean on Callisto. *Nature* 412: 409–411.
Ruiz J (2003) Heat flow and depth to a possible internal ocean on Triton. *Icarus* 166: 436–439.
Sandwell D and Schubert G (2010) A contraction model for the flattening and equatorial ridge of Iapetus. *Icarus* 210: 817–822.
Schenk PM and McKinnon WB (2009) One-hundred-km-scale basins on Enceladus: Evidence for an active ice shell. *Geophysical Research Letters* 36: L16202.
Schenk PM and Turtle EP (2009) Europa's impact craters: Probes of the icy shell. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 181–198. Tucson, AZ: University of Arizona Press.
Schenk PM and Zahnle K (2007) On the negligible surface age of Triton. *Icarus* 192: 135–149.
Schmidt BE, Blankenship DD, Patterson GW, and Schenk PM (2011) Active formation of 'chaos terrain' over shallow subsurface water on Europa. *Nature* 479: 502–505.
Schmidt J, Brilliantov N, Spahn F, and Kempf S (2008) Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. *Nature* 451: 685–688.
Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure, and dynamics of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter. The Planet, Satellites, and Magnetosphere*, pp. 281–306. Cambridge: Cambridge University Press.
Schubert G, Anderson JD, Travis BJ, and Palguta J (2007) Enceladus: Present internal structure and differentiation by early and long-term radiogenic heating. *Icarus* 188: 345–355.
Schubert G, Cassen P, and Young RE (1979) Subsolidus convective cooling histories of terrestrial planets. *Icarus* 38: 192–211.

Schubert G, Sohl F, and Hussmann H (2009) Interior of Europa. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 353–368. Tucson, AZ: University of Arizona Press.

Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories, compositions and internal structures of the moons of the solar system. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 224–292. Tucson, AZ: Arizona University Press.

Schubert G, Stevenson DJ, and Ellsworth K (1981) Internal structures of the Galilean satellites. *Icarus* 47: 46–59.

Schubert G, Zhang K, Kivelson MG, and Anderson JD (1996) The magnetic field and internal structure of Ganymede. *Nature* 384: 544–545.

Segatz M, Spohn T, Ross MN, and Schubert G (1988) Tidal dissipation, surface heat flow and figures of viscoelastic models of Io. *Icarus* 75: 187–206.

Showman AP and Han L (2004) Numerical simulation of convection in Europa's ice shell: Implications for surface features. *Journal of Geophysical Research* 109(E1). http://dx.doi.org/10.1029/2003JE002103.

Showman AP and Malhotra R (1997) Tidal evolution into the Laplace resonance and the resurfacing of Ganymede. *Icarus* 127: 93–111.

Showman AP, Stevenson DJ, and Malhotra R (1997) Coupled orbital and thermal evolution of Ganymede. *Icarus* 129: 367–383.

Singer KN and McKinnon WB (2011) Tectonics on Iapetus: Despinning, respinning or something completely different? *Icarus* 216: 198–211.

Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.

Soderblom LA, Becker TL, Kieffer SW, Brown RH, Hansen CJ, and Johnson TV (1990) Triton's geyser-like plumes – Discovery and basic characterization. *Science* 250: 410–415.

Sohl F, Hussmann H, Schwentker B, and Spohn T (2003) Interior structure and tidal Love numbers of Titan. *Journal of Geophysical Research* 108(E12): 5130. http://dx.doi.org/10.1029/2003JE002044.

Sohl F, Sears WD, and Lorenz R (1995) Tidal dissipation on Titan. *Icarus* 115: 278–294.

Sohl F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo observations on the interior structure and chemistry of the Galilean satellites. *Icarus* 157: 104–119.

Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.

Solomatov VS and Moresi L-N (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on Earth and other terrestrial planets. *Journal of Geophysical Research* 105(B9): 21795–21818.

Sotin C, Grasset O, and Beauchesne S (1998) Thermodynamical properties of high pressure ices: Implications for the dynamics and internal structure of large icy satellites. In: Schmitt B, et al. (eds.) *Solar System Ices*, pp. 79–96. Norwell, MA: Kluwer.

Sotin C, Head JW, and Tobie G (2002) Europa: Tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting. *Geophysical Research Letters* 29(8): 74-1–74-4.

Sotin C, Jaumann R, Buratti BJ, et al. (2005) Release of volatiles from a possible cryovolcano from near-infrared imaging of Titan. *Nature* 435: 786–789.

Sotin C and Tobie G (2004) Internal structure and dynamics of the large icy satellites. *Comptes Rendus De L'Academie Des Sciences Physique* 5: 769–780.

Sotin C, Tobie G, Wahr J, and McKinnon WB (2009) Tides and tidal heating on Europa. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 85–118. Tucson, AZ: University of Arizona Press.

Spahn F, Schmidt J, Albers N, et al. (2006) Cassini dust measurements at Enceladus and implications for the origin of the E-ring. *Science* 311: 1416–1418.

Spaun NA, Head JW, Collins GC, Prockter LM, and Pappalardo RT (1998) Conamara Chaos region, Europa: Reconstruction of mobile polygonal ice blocks. *Geophysical Research Letters* 25: 4277–4280.

Spencer JR, Barr AC, Esposito LW, et al. (2009) Enceladus: An active cryovolcanic satellite. In: Dougherty MK, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 683–724. Dordrecht: Springer.

Spencer JR and Nimmo F (2013) Enceladus: An active ice world in the Saturn system. *Annual Review of Earth and Planetary Sciences* 41: 693–717.

Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. *Science* 311: 1401–1405.

Spitale JN and Porco CC (2007) Association of the jets on Enceladus with the warmest regions on its south-polar fractures. *Nature* 449: 695–697.

Spohn T (1997) Tides of Io. In: Wilhelm H, Zürn W, and Wenzel H-G (eds.) *Tidal Phenomena*, pp. 345–377. Berlin: Springer.

Spohn T and Schubert G (2003) Oceans in the icy Galilean satellites of Jupiter? *Icarus* 161: 456–467.

Squyres SW, Reynolds RT, and Cassen PM (1983a) Liquid water and active resurfacing on Europa. *Nature* 301: 225–226.

Squyres SW, Reynolds RT, Cassen PM, and Peale SJ (1983b) The evolution of Enceladus. *Icarus* 53: 319–331.

Stern SA and McKinnon WB (2000) Triton's surface age and impactor population revisited in light of Kuiper belt fluxes: Evidence for small Kuiper belt objects and recent geological activity. *The Astronomical Journal* 119: 945–952.

Stevenson DJ (1992) *The Interior of Titan*. Proceedings Symposium on Titan, ESA SP-338, pp. 29–33. Noordwijk, The Netherlands: ESA.

Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.

Stevenson DJ and Lunine JI (1988) Rapid formation of Jupiter by diffuse redistribution of water vapour in the solar nebula. *Icarus* 75: 146–155.

Stofan ER, Lunine JI, Lopes R, et al. (2006) Mapping of Titan: Results from the first Titan radar passes. *Icarus* 185: 443–456.

Strobel DF (1974) The photochemistry in the atmosphere of Titan. *Icarus* 21: 466–470.

Tholen DJ and Buie MW (1990) Further analysis of the Pluto–Charon mutual event observations. *Bulletin of the American Astronomical Society* 22: 1129.

Thomas PC (2000) The shape of Triton from limb profiles. *Icarus* 148: 587–588.

Thomas PC (2010) Sizes, shapes, and derived properties of the Saturnian satellites after the Cassini nominal mission. *Icarus* 208: 395–401.

Thomas PC, Burns JA, Helfenstein P, et al. (2007) Shapes of the Saturnian icy satellites and their significance. *Icarus* 190: 573–584.

Thomson RE and Delaney JR (2001) Evidence for a weakly stratified European ocean sustained by seafloor heat flux. *Journal of Geophysical Research* 106: 12355–12365.

Tobie G, Cadek O, and Sotin C (2008) Solid tidal friction above a liquid water reservoir as the origin of the south pole hotspot on Enceladus. *Icarus* 196: 642–652.

Tobie G, Choblet G, and Sotin C (2003) Tidally heated convection: Constraints on Europa's ice shell thickness. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2003JE002099.

Tobie G, Lunine J, and Sotin C (2006) Episodic outgassing as the source of atmospheric methane on Titan. *Nature* 440: 61–64. http://dx.doi.org/10.1038/nature04497.

Tobie G, Mocquet A, and Sotin C (2005) Tidal dissipation within large icy satellites: Europa and Titan. *Icarus* 177: 534–549.

Tomasko M, Archinal B, Becker T, et al. (2005) Rain, winds, and haze during the Huygens probe descent to Titan's surface. *Nature* 438: 765–778.

Travis BJ, Palguta J, and Schubert G (2012) A whole-moon thermal history model of Europa: Impact of hydrothermal circulation and salt transport. *Icarus* 218: 1006–1019.

Tyler RH (2008) Strong ocean tidal flow and heating of moons of the outer planets. *Nature* 456: 770–772.

Tyler GL, Sweetnam DN, Anderson JD, et al. (1989) Voyager radio science observations of Neptune and Triton. *Science* 246: 1466–1473.

Vance S and Goodman J (2009) Oceanography of an ice-covered moon. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 459–484. Tucson, AZ: University of Arizona Press.

Van Hoolst T, Baland R-M, and Trinh A (2013) On the librations and tides of large icy satellites. *Icarus* 226: 299–315.

Van Hoolst T, Rambaux N, Karatekin Ö, Dehant V, and Rivoldini A (2008) The librations, shape and icy shell of Europa. *Icarus* 195: 386–399.

Wahr JM, Zuber MT, Smith DE, and Lunine JI (2006) Tides on Europa and the thickness of Europa's icy shell. *Journal of Geophysical Research* 111: E12. http://dx.doi.org/10.1029/2006JE002729.

Waite JH, Combi MR, Ip W-H, et al. (2006) Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure. *Science* 311: 1419–1422.

Waite JH Jr., Lewis WS, Magee BA, et al. (2009) Liquid water on Enceladus from observations of ammonia and ^{40}Ar in the plume. *Nature* 460: 487–490.

Weertman J (1983) Creep deformation of ice. *Annual Review of Earth and Planetary Sciences* 11: 215–240.

White OL, Schenk PM, and Dombard AJ (2013) Impact basin relaxation on Rhea and Iapetus and relation to past heat flow. *Icarus* 223: 699–709.

Wisdom J (2004) Spin-orbit secondary resonance dynamics of Enceladus. *The Astronomical Journal* 128: 484–491.

Wong M, Lunine JI, Atreya SK, et al. (2008) Oxygen and other volatiles in the giant planets and their satellites. *Reviews in Mineralogy and Geochemistry* 68: 219–246.

Yoder CF (1979) How tidal heating in Io drives the Galilean orbital resonance locks. *Nature* 279: 767–770.

Yoder CF and Peale SJ (1981) The tides of Io. *Icarus* 47: 1–35.

Zahnle K, Schenk P, Levison H, and Dones L (2003) Cratering rates in the outer solar system. *Icarus* 163: 263–289.

Zharkov VN, Leontjev VV, and Kozenko AV (1985) Models, figures, and gravitational moments of the Galilean satellites of Jupiter and icy satellites of Saturn. *Icarus* 61: 92–100.

Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.

Zolotov MY and Kargel JS (2009) On the chemical composition of Europa's icy shell, ocean, and underlying rocks. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 431–445. Tucson, AZ: University of Arizona Press.

Zschau J (1978) Tidal friction in the solid Earth: Loading tides versus body tides. In: Brosche P and Sündermann J (eds.) *Tidal Friction and the Earth's Rotation*, pp. 62–94. Berlin: Springer.

This page intentionally left blank

10.19 Introduction to 'Pluto, Charon, and the Kuiper Belt Objects': Pluto on the Eve of the *New Horizons* Encounter

WB McKinnon, Washington University in St Louis, Saint Louis, MO, USA

© 2015 Elsevier B.V. All rights reserved.

10.19.1	**Introduction**	637
10.19.2	**The Pluto–Charon System**	637
10.19.2.1	Surface Compositions of Pluto and Charon	638
10.19.2.2	Pluto's Atmosphere	640
10.19.2.3	Interior and Geologic Activity	641
10.19.3	**The Largest Trans-Neptunian Dwarf Planets and Dwarf Planet Candidates**	643
10.19.3.1	Surface Compositions	643
10.19.3.2	Densities and Interiors	644
10.19.4	**The Kuiper Belt**	645
10.19.4.1	Nice Models	646
10.19.4.2	Kuiper Belt Population Characteristics	648
10.19.5	**Final Remarks**	648
Acknowledgments		648
References		649

10.19.1 Introduction

Approach-phase operations began in January 2015 for the *New Horizons* spacecraft, in advance of its planned fly-through of the Pluto–Charon system on 14 July 2015 (Stern, 2008; Weaver et al., 2008; Young et al., 2008b). As the second edition of the *Treatise on Geophysics* was being prepared before the encounter, these anticipated results could not be incorporated. Nonetheless, in the past several years a substantial amount of new information has been gleaned by ground-based and space-based study of Pluto and its moons and especially of the Kuiper Belt (more broadly, the entire trans-Neptunian region). Theoretical developments in the case of the Kuiper Belt as a whole have also been notable and have driven an entirely new understanding of the evolution of our solar system.

With the agreement of the editors and original authors, what follows is a brief introduction to Pluto and the Kuiper Belt as of early 2014, updating some key aspects of the original article (which follows) and offering a prospectus for what may be learned from *New Horizons*. In outline, I will present (1) an updated view of the Pluto–Charon system, (2) a summary of salient aspects of Pluto's fellow dwarf planets and related trans-Neptunian objects (TNOs), and (3) a very brief overview of the structure of the Kuiper Belt and its implications. While limited in scope and necessarily selective, it is hoped that this brief review will serve as a springboard to dive into the primary literature of this fast-moving field.

10.19.2 The Pluto–Charon System

The most obvious change since the first edition of the *Treatise* is that two additional satellites have been discovered, now named Styx and Kerberos (**Table 1**). These are rather small moons in comparison with either Pluto or Charon and required the limiting magnitude capabilities of the *Hubble Space Telescope* (*HST*) to be detected at all (Showalter et al., 2011, 2012). It is doubtful that our knowledge of the architecture of the Pluto–Charon system is complete, and it is anticipated that *New Horizons* will discover additional moons on approach (in addition to their intrinsic scientific interest, there is the practical matter of identifying any potential dust or debris hazards to the flyby associated with such moons or moonlets).

It is notable that the entire Pluto–Charon system is coplanar and that the orbital eccentricities and inclinations are very low (or zero!) (Buie et al., 2012b). This implies that the system is highly dynamically relaxed and that the small satellites (at least) formed from a dissipative, presumably particulate, disk. That Pluto and Charon represent a binary pair that has achieved mutual spin–orbit synchronism has been appreciated for some time, and Charon is thought to have tidally evolved outward from an orbit much closer to Pluto, following that moon's formation in a massive collision ('giant impact') (e.g., Dobrovolskis et al., 1997). Similarly, the accretion of Pluto's small moons from an extended particle disk, by standard, gravitational coagulation processes, has been shown to be physically plausible (Bromley and Kenyon, 2014). The trick, however, is to make such an extended disk in the first place. Disk formation presumably also occurred during the giant impact that formed Charon, but a sufficiently *extended* disk has not been an outcome in numerical, Charon-forming impact models to date (Canup, 2005, 2011).

Alternatively, Pluto's small moons formed much closer in, and Charon tidally captured them as it evolved away from Pluto, subsequently driving them all, in lockstep, outward (e.g., Stern et al., 2006). While appealing at first, **Chapter 10.17** makes clear that this explanation suffers (what appear to be) insurmountable physical difficulties. First, the small satellites as a whole are not obviously in resonance today (although the orbital periods of Styx, Nix, Kerberos, and

Table 1 Properties of the Pluto–Charon system (Brozović et al., 2014)

	Semimajor axis[a] *(km)*	*Period*[a] *(days)*	*Eccentricity*[a]	*Inclination*[a] *(°)*	*Radius*[b] *(km)*	*Density*[b] *(g cm^{-3})*	*Albedo*[c] *(%)*
Pluto		6.3872			1169–1193	1.89 ± 0.06	49–66
Charon	19 596	6.3872	0.00005	0.0	603.6 ± 1.4	1.72 ± 0.02	36–39
Styx	42 413	20.1617	0.00001	0.0	4–14	$0.0^{+5.0}_{-0.0}$	4–35
Nix	48 690	24.8548	0.00000	0.0	23–70	$0.1^{+0.2}_{-0.1}$	4–35
Kerberos	57 750	32.1679	0.00000	0.4	7–22	$1.3^{+2.2}_{-1.3}$	4–35
Hydra	64 721	38.2021	0.00554	0.3	29–86	$0.1^{+0.1}_{-0.1}$	4–35

[a]Mean orbital elements for Charon are plutocentric, whereas those for the small satellites are barycentric.
[b]Formal uncertainties are 1σ.
[c]Geometric albedos for Pluto and Charon vary with location, and especially strongly so for Pluto (Buie et al., 2010b); the ranges quoted are hemispheric averages (Buie et al., 1997). The albedo range for the small satellites is assumed and determines the radius ranges shown.

Hydra are, respectively, close to the 3:1. 4:1, 5:1, and 6:1 mean-motion resonances with Charon). Second, all resonant evolution scenarios lead to strong growth in one or more small satellite eccentricities, eccentricities that are not hinted at today (**Table 1**). The reader is directed to **Chapter 10.17** in this volume for a full discussion of this matter. Further clues to the formation and evolution of the Pluto–Charon system will no doubt be uncovered during the *New Horizons* encounter.

A by-product of the discovery of the small satellites and the detailed study of their orbits has been a relatively precise determination of the system barycenter and thus the Pluto/Charon mass ratio. Coupled with precise determinations of Charon's size by multiple stellar occultations (Person et al., 2006; Sicardy et al., 2006, 2012), Charon's mean density is now relatively well determined, both in an absolute sense and with respect to Pluto. Pluto's precise size remains uncertain due to the effects of its atmosphere on stellar occultation interpretations (**Section 10.19.2.2**) and limb darkening effects on mutual-event modeling (referring to the 1985–1990 occultations and transits of Pluto and Charon as seen from the Earth (Binzel and Hubbard, 1997; Tholen and Buie, 1997)). Both Pluto and Charon are relatively rock-rich, but Charon is somewhat icier, which is consistent with impact origin scenarios in which one or both precursor bodies were at least partially differentiated (rock separated from ice under the action of gravity) (Canup, 2005; see McKinnon et al., 2008).

Ice-rich ejecta is also the preferred route for forming a particle disk outside of Charon's initial post-giant impact orbit (Canup, 2011). In this regard, the limits on the densities of the small satellites in **Table 1** are intriguing. Although the lower limits are consistent with zero (no mass detected), the upper limits on the densities of the brightest and presumably largest of the small satellites (Nix and Hydra) are – even at the 3σ error level – consistent with very icy and porous bodies. Icy rubble-pile bodies are in fact predicted by the initial conditions of the giant impact as presently understood (conditions necessary to give an icier Charon). The small satellites will thus offer an interesting cosmogonic test during the *New Horizons* encounter, as good constraints will be placed on satellite volumes from imaging and improved satellite masses will result from optical navigation and orbital integrations (noting *New Horizons* will not pass close enough to any of the small satellites to measure their gravitational accelerations directly).

10.19.2.1 Surface Compositions of Pluto and Charon

Understanding of the basic composition of Pluto's surface has not fundamentally changed in the last few years, but there have been important refinements. Pluto's visible and near-infrared spectra are dominated by methane (CH_4) absorptions, but much of the optical surface is physically dominated by nitrogen (N_2) ice in which the methane (along with some CO) is dissolved as a minor component in solid solution (Cruikshank et al., 1997). (As is well known, methane absorbs strongly in the near-infrared and tends to dominate the spectra of bodies on which CH_4 is present, even in minor amounts.) Methane-rich surface units also exist (Douté et al., 1999; Olkin et al., 2007), which likely couple directly to the atmosphere, and earlier detections of solid ethane (C_2H_6) have been essentially confirmed, at the few percent level as a pure solid (DeMeo et al., 2010; Merlin et al., 2010).

HST images have also been used to create the best maps of the surface albedo and color distribution to date (Buie et al., 2010b). Data were taken at two visible wavelengths and show widespread albedo and color variegation on the surface of Pluto (**Figure 1**). As can be seen from the figure, the mid- to high-latitude region on Pluto's sunlit pole is, on average, more neutral in color and of generally higher albedo than the rest of the surface. Presumably, the polar surface, which has emerged from its seasonal winter over the last few decades, is rich in nitrogen ice (and its dissolved constituents). The equatorial bright spot near 180° longitude is also dominated by nitrogen ice and is relatively CO ice-rich (Grundy and Buie, 2001; Grundy et al., 2013).

The darker regions on Pluto show considerable color diversity, arguing that there must be a range of compositional units in the dark regions. The methane-ice-rich surface units mentioned earlier (which are, paradoxically, indicated by intrinsically weak CH_4 bands) are not correlated with the brighter and more spectrally neutral surface units in **Figure 1**, but neither do maximum CH_4 absorptions (of any intrinsic strength) correlate with the darkest region (and light-curve minimum) near 80–120° longitude (Grundy et al., 2013). The lowest geometric albedos on Pluto (a few %) occur in this expanse, and low albedos on Pluto generally are thought to be due to long-term photolytic and radiolytic processing of methane ice into dark, reddish, carbonaceous heteropolymers (or *tholins*) and

Table 2 Ten largest trans-Neptunian objects, plus Ceres

Object	*Class*[a]	$\langle a \rangle$[b] *(AU)*	$\langle e \rangle$[b]	$\langle i \rangle$[b] *(°)*	r_{2020}[c] *(AU)*	*Radius*[d] *(km)*	*Reported satellites*	*Known Surface ices*[e]
Pluto	3:2	39.5	0.24	16	34	1180	5	CH_4, N_2, CO, C_2H_6
Charon	〃	〃	〃	〃	〃	605		H_2O, NH_3
Eris	SDO/DO	67.9	0.45	43	96	1160	1	CH_4+
Makemake	HC	45.6	0.16	28	53	715		CH_4+, C_2H_6
2007 OR_{10}	10:3	67.2	0.45	35	88	640		H_2O, CH_4?
Haumea	HC	43.1	0.21	26	50	620	2	H_2O
Quaoar	CC	43.4	0.04	9	43	555	1	H_2O, CH_4, C_2H_6
Sedna	DO	528	0.86	12	84	500		H_2O? CH_4?
Ceres	MBA	2.77	0.116	9.6	2.97	475		
2002 MS_4	HC	41.8	0.15	19	46	465		
Orcus	3:2	39.3	0.23	21	48	460	1	H_2O, NH_3?

[a]SDO = scattered disk object, CC = cold classical, HC = hot classical, and DO = detached object (following Gladman et al., 2008); 3:2 and 10:3 refer to mean-motion resonances with Neptune; Ceres is a main belt asteroid (MBA).

[b]$\langle a \rangle$, $\langle e \rangle$, and $\langle i \rangle$ are mean semimajor axis, eccentricity, and inclination with respect to the invariable plane over a 10 My integration (Elliot et al., 2005; www.boulder.swri.edu/~buie/kbo/astrom), except Ceres (proper elements from hamilton.dm.unipi.it/astdys).

[c]r_{2020} is distance from the Sun during 2020 (01 July for Ceres).

[d]Radius estimates from occultations and analyses of *Spitzer* (Stansberry et al., 2008) and *Herschel* (Fornasier et al., 2013; Lim et al., 2010; Pál et al., 2012; Santos-Sanz et al., 2012; Vilenius et al., 2012) observations.

[e]'CH_4+' indicates a contaminant, possibly N_2, implied by CH_4 band shifts. CH_4 is suspected on 2007 OR_{10} (Brown et al., 2011), and CH_4 and H_2O ice could be on Sedna (Barucci et al., 2010; Emery et al., 2007).

Source: Grundy WM, McKinnon WB, Ammannito E, et al. (2009) *Exploration Strategy for the Ice Dwarf Planets 2013–2022*. SBAG Community White Paper for the Planetary Science Decadal Survey. Washington, DC: National Academy of Sciences.

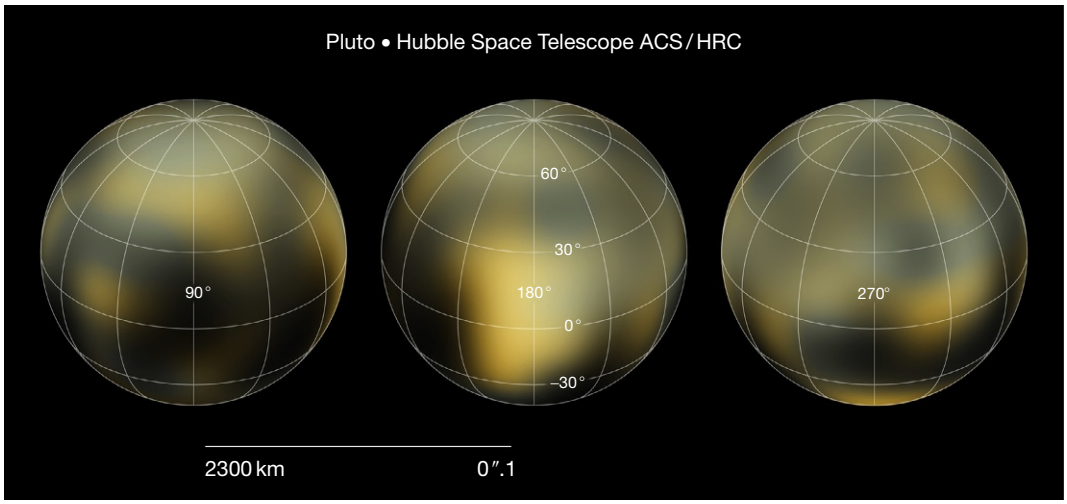


Figure 1 Albedo and color maps of Pluto created from *HST* Advanced Camera for Surveys/High Resolution Camera (ACS/HRC) images taken at 0.435 and 0.555 μm (based on Buie et al., 2010b); 0″.1 refers to angular size as seen from the Earth. Pluto's presently illuminated, south (in a celestial sense) pole is up, but as Pluto is a retrograde rotator, longitude increases to the 'east' as defined by the planet's spin. These maps will be superseded in resolution in mid 2015, by *New Horizons* imaging as the spacecraft approaches Pluto. NASA/ESA/M. Buie (SwRI).

ultimately into dark, spectrally neutral carbon (Cruikshank et al., 1997; de Bergh et al., 2008; Hudson et al., 2008; Olkin et al., 2007).

It is notable that *no* solid CO_2 has been detected so far on the surface of Pluto (Grundy et al., 2013), whereas near-infrared absorptions for CO_2 ice are quite pronounced on Triton. Triton is Neptune's large, retrograde moon and likely a captured Kuiper Belt object (KBO) and cousin to Pluto (McKinnon et al., 1995); the CO_2 absorptions on Triton imply substantial exposures of this relatively involatile 'bedrock' on this satellite (Grundy et al., 2010; Quirico et al., 1999). The spectral signature of water ice is also seen on Triton and not seen on Pluto, and H_2O ice is either unnecessary or contraindicated when modeling Pluto's spectra (cf. DeMeo et al., 2010; Grundy et al., 2013; Merlin et al., 2010; Olkin et al., 2007). It may be that the pronounced surface–atmosphere interactions on Pluto have buried, coated, or otherwise obscured exposures of CO_2 or H_2O ice on Pluto, or such exposures may be very limited; it may also be that the more volatile ices (N_2 and CH_4) form a geologically substantial

crustal layer on Pluto. Imaging spectroscopy from *New Horizons* should prove telling, especially if the volatile ices are thin enough that deeper materials are exposed by impacts or other geologically disruptive phenomena (Moore et al., 2014).

Charon's near-infrared spectrum, in contrast, is *dominated* by the signature of crystalline water ice (H_2O), with no hint of N_2, CH_4, CO, or CO_2 (e.g., Merlin et al., 2010). Remarkably, however, it is the spectral signature of ammonia hydrate ice ($NH_3{\cdot}nH_2O$, in some form or forms) that is seen on Charon, near 2.2 μm (Brown and Calvin, 1999; Cook et al., 2007; Dumas et al., 2001; Merlin et al., 2010; **Figure 2**). Ammonia, due to its spectacular effect on lowering the minimum melting point (or eutectic) of ammonia hydrate (e.g., Kargel, 1992), is long thought to have been involved in 'cryovolcanic' activity in the outer solar system, but had not been confirmed on a solid solar system body until the detections on Charon. Ammonia is thought to be the ultimate source for much or all of the N_2 in the atmosphere of Titan (Atreya et al., 2009) and perhaps on large KBOs such as Pluto as well (McKinnon et al., 1997). Ammonia is, of course, known as a component of cometary ices (Mumma and Charnley, 2011) and is seen in the vapor component of the Enceladus plume (Waite et al., 2009), in both cases at the percent level or less compared with H_2O.

Ammonia, whether hydrated or not, is notoriously unstable under UV and charged particle irradiation (Lanzerotti et al., 1984; Strazzulla and Palumbo, 1998), so it is important to consider what makes Charon special in this regard. Is it simply its distance from the Sun, or does Pluto's escaping atmosphere have something to do with it, perhaps by protecting Charon from direct solar wind bombardment? Cook et al. (2007) proposed that we see ammonia hydrate on Charon because of recent cryovolcanic activity, while Dumas et al. (2001) speculated that the ammonia hydrate is a product of implantation of N_2/N^+ from Pluto's atmosphere into Charon's water ice surface, followed by radiolytic processing. Perhaps something as prosaic as impact exhumation of subsurface ammonia hydrate is involved. Whatever the situation, *New Horizons* imaging and spectral capabilities should provide important new constraints on the origin of ammonia on Charon and perhaps other large KBOs (see **Section 10.19.3.1**).

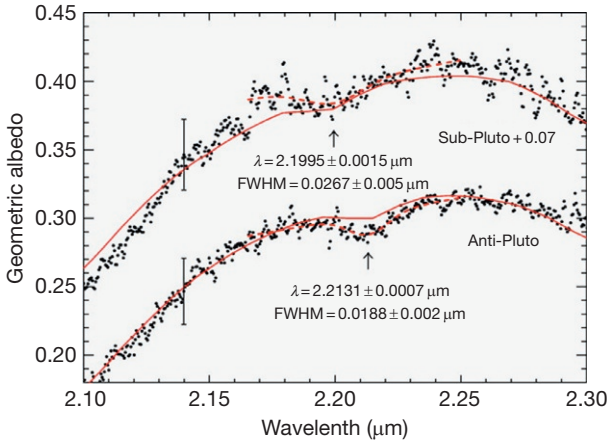


Figure 2 Ammonia hydrate feature in the spectrum of Charon. Solid red lines indicate a model fit (water ice, continuum absorber, plus ammonia hydrate); dashed lines are a Gaussian fit to the absorption feature. The absorption feature is seen at all aspects: on the sub- and anti-Pluto hemispheres (as shown) and on the leading and trailing hemispheres of motion (Verbiscer et al., 2007). Reproduced from Cook JC, Desch SJ, Roush TL, Trujillo CA, and Geballe TR (2007) Near infrared spectroscopy of Charon: Possible evidence for cryovolcanism on Kuiper belt objects. *Astrophysical Journal* 663: 1406–1419.

10.19.2.2 Pluto's Atmosphere

Pluto's atmosphere is supported by the vapor pressure of nitrogen frost. Stellar occultations offer a sensitive probe of that atmosphere and thus of the temperature of Pluto's surface nitrogen ice and the seasonal evolution of the surface–atmosphere couple. The last few years have a boon to occultation studies as Pluto (as seen from the Earth) traversed the galactic plane, leading to observed occultations in 2006 (Elliot et al., 2007; Young et al., 2008a), 2007 (McCarthy et al., 2008; Person et al., 2008), 2008 (Sicardy et al., 2011), 2010 (Person et al., 2010), 2011 (Person et al., 2013), 2012 (Bosh et al., 2013), and 2013 (Olkin et al., 2014). A typical, derived temperature profile is shown in **Figure 3(a)**, along with various relevant model aspects or benchmarks. The retrieved thermal profile displays a warm (~100 K) upper atmosphere and a sharp, stratospheric temperature inversion (temperature increasing with height), as have atmospheric thermal profiles since Pluto's atmosphere was first discovered, during a stellar occultation (Elliot et al., 1989). A photochemical or condensation haze layer may also be present and was one explanation for the pronounced, symmetrical 'kink' or 'knee' in the original occultation light curve from 1988. The knee was greatly diminished or absent in light curves from later years but has apparently returned or strengthened in 2011–2013 (Olkin et al., 2014; Person et al., 2013). This is likely a testament to the variability of Pluto's atmosphere on decadal timescales, even if the physical explanation is unclear.

Pluto's atmosphere expanded and increased in mass (and thus surface pressure) by a factor of 2 between the 1988 and 2002 occultations (Elliot et al., 2003, 2007). During this time, the planet passed through perihelion (peak insolation) *and* equinox, which turns out to be important for understanding this behavior (Young, 2013). The inference of increased atmospheric mass is now backed up by general circulation models (Zalucha and Gulbis, 2012). Later occultations (2006 through 2013) show that this 'bulked up' atmosphere has been maintained since 2002 (e.g., Young et al., 2008a) and the atmosphere may even be increasing in mass at present (Olkin et al., 2014). Pluto's atmosphere will thus be fully available for study by *New Horizons* and, in the favored model of Olkin et al., does not completely collapse at all during the Pluto year, because of the quantity of nitrogen ice available at both poles and the high thermal inertia of the ice assumed in the favored model. Pluto's atmospheric variations are presumably tied to the surface albedo, color, and spectral variations observed on the same time frame (Buie et al., 2010a,b; Grundy et al., 2014).

The decrease in atmospheric temperature with height in Pluto's upper atmosphere in **Figure 3(a)** is real. It is a product of both radiative cooling by CO and the expansion of Pluto's atmosphere (via enhanced Jeans escape) out into space (Zhu et al., 2014). The temperature maximum is due to heating due to infrared absorption by CH_4 (Yelle and Lunine, 1989). Both

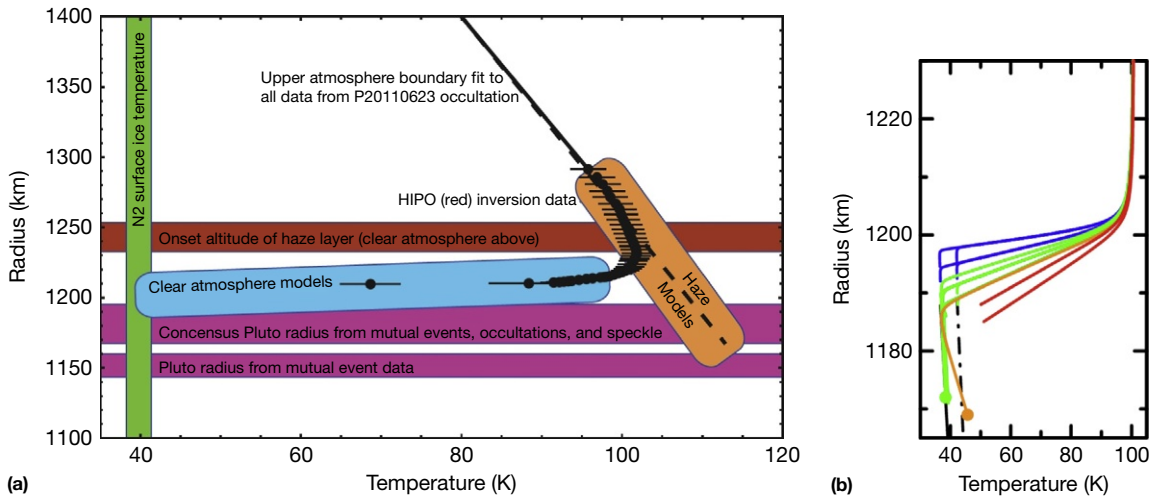


Figure 3 (a) Derived temperature profile from the 23 June 2011 stellar occultation of Pluto's atmosphere, along with model elements and constraints. Reproduced from Person MJ, Dunham EW, Bosh AS, et al. (2013) The 23 June 2011 stellar occultation by Pluto: airborne and ground observations. *Astronomical Journal* 146:83 (15 pp.). HIPO refers to the High Speed Imaging Photometer for Occultations aboard *SOFIA*. Models with haze extend warm 'thermospheric' temperatures downward, but the profile must eventually turn back toward the cold surface. (b) Family of possible temperature profiles for Pluto's atmosphere. Modified from Lellouch E, Sicardy B, de Bergh C, Käufl H-U, Kassi S, and Campargue A (2009) Pluto's lower atmosphere structure and methane abundance from high-resolution spectroscopy and stellar occultations. *Astronomy & Astrophysics* 495: L17–L21. All of the models in b assume a clear atmosphere (no haze) with different stratospheric thermal gradients. Most assume little or no cool, basal troposphere. Solid black vertical line is the N_2 saturation curve; dot-dashed line is saturation curve for a 0.5% CH_4 mixing ratio. Tropospheric temperature gradients are shown following either the wet (green) or dry (orange) N_2 adiabat, with maximum tropospheric depths indicated by dots. Reproduced with permission from *Astronomy & Astrophysics*, © ESO.

gaseous CH_4 and CO have now been directly detected in Pluto's atmosphere (Lellouch et al., 2009, 2011) (in addition to being known components of Pluto's surface ices), which supports theoretical models of the atmosphere (Yelle and Elliot, 1997). The expansion and escape of Pluto's atmosphere are driven by far- and extreme-ultraviolet heating, and the exobase is predicted to be out near 8 Pluto radii at present (Zhu et al., 2014). The *New Horizons* trajectory may thus take the spacecraft directly through the planet's extended atmosphere (depending, of course, on the atmosphere's interaction with the solar wind at encounter time).

Due to differential atmospheric refraction, Pluto's solid surface cannot be seen in stellar occultations in the present epoch, from the Earth. Light curves probe down to the ≈1190–1200 km radius level at best (Lellouch et al., 2014; Young et al., 2008a). **Figure 3(b)** illustrates possible Pluto temperature profiles for a range of stratospheric temperature gradients that are constrained by occultation light curves *and* the relatively warm temperature of the methane gas detected by Lellouch et al. (2009). This limits the possible depths of cold tropospheres as shown. Lellouch et al. (2009, 2014) argued that any troposphere on Pluto may be limited to a scale height (20 km) or less, which is consistent with the view of emerging general circulation models (Zalucha et al., 2011). As **Figure 3(a)** alludes to, there has been a long-standing mismatch between estimates of Pluto's size from mutual-event data and stellar occultations.

Matters pertaining to Pluto's atmosphere will be vastly clarified by the *New Horizons* encounter in 2015. Pluto's solid-body radius will be determined once and for all by imaging, and the distribution of surface ices and their compositions will be mapped by the infrared spectrometer. Surface temperatures will be measured by radio science (antenna and receiver) operating in a passive mode. UV solar and stellar occultations will determine the atmosphere's composition and thermal structure in its upper reaches, while radio science will sound the pressure and temperature of the lower atmosphere. All these data should provide an exquisite picture of Pluto's surface and atmosphere at a single epoch, one that can be used as a tie point to interpret both older measurements and new data to come as Pluto sails off to aphelion.

10.19.2.3 Interior and Geologic Activity

Among the most anticipated aspects of the *New Horizons* flyby are detailed images and spectra of the surfaces of Pluto and Charon (resolutions exceeding 100 m px^{-1} for Pluto, for the nominal trajectory). These new data will quantify the level of geologic activity on these bodies, now and in the past, and in Pluto's case should provide an important comparison with Triton, which was visited by Voyager 2 in 1989 (Moore et al., 2014).

The fly-through of the Pluto–Charon system by *New Horizons* also offers the first opportunity to determine both the total system mass and the individual masses of Pluto and Charon by direct Doppler tracking. Simulations of the encounter by the Radio Science Experiment (REX) indicate the potential accuracies of the combined and individual mass determinations of Pluto and Charon are on the order of 0.1% (Paetzold et al., 2013; Tyler et al., 2008), a substantial improvement in precision (especially for Charon). This will put to rest a long-standing limitation on modeling and understanding Pluto and Charon's internal structure and evolution (McKinnon et al., 1997).

Based on the values in **Table 1**, Pluto's rock/ice ratio is 0.65–0.70 by mass, calculated on an anhydrous basis and neglecting the planet's bulk carbonaceous component. Pluto's

differentiation state was debated in McKinnon et al. (1997), but the case for Pluto being differentiated (having a rocky core) has strengthened considerably since then. First, solid-state convection is now seen as a less efficient mechanism of interior heat transport in the so-called stagnant lid regime, in which a cold, conductive, and essentially immobile lid overlies a much warmer, convecting interior. This regime, which should apply to Pluto and Charon, has been shown to imply greater internal temperatures than had been estimated by earlier theory (see, e.g., McKinnon, 2006; Solomatov and Moresi, 2000; **Chapter 10.08** in this volume, and the comprehensive overview of Schubert et al., 2001). This implies greater internal temperatures, all other things being equal, and if internal temperatures are higher, the likelihood of ice melting and thus the inward settling of denser rock under the action of gravity (differentiation) are greater. Second, Charon's somewhat lower density than Pluto implies at least partial differentiation of one of the precursor bodies, prior to the Charon-forming collision (as noted in **Section 10.19.2.1**). Third, at least one other major KBO, Haumea, is clearly differentiated (Brown, 2008) (see **Section 10.19.3.1**). If differentiated, Pluto should have a relatively large rocky core, approaching 900 km radius, depending on core rock hydration state and the extent of any refractory carbonaceous component.

It is also possible, from the densities in **Table 1**, that Pluto and Charon have the same or similar densities, considering the stated uncertainties at the 3σ level and the effects of porosity (which should be greater for Charon). If Pluto and Charon were close in density after all, it could change our view as to how the binary formed. The precursor bodies, or at least the impactor, would then have to have been undifferentiated (Canup, 2005). Similar densities would also allow consideration of formation of the Pluto–Charon binary by direct gravitational instability (Nesvorný et al., 2010), a formation mechanism presently disallowed because it does not predict manifestly different densities or compositions for individual binary members. On the other hand, mutual-event solutions permit a substantially denser Pluto than those given in **Table 1** (Tholen and Buie, 1997). Determining the precise Pluto/Charon *density* ratio is clearly an important objective for *New Horizons*.

Exploration of the outer solar system has also revealed that many large icy satellites possess or probably possess interior oceans, among them Europa, Ganymede, Callisto, and Titan, and even midsize satellites (i.e., Enceladus) may possess oceans or seas if they are tidally heated (see Sohl et al. (2010) and **Chapter 10.18** in this volume for detailed reviews). The efficiency of solid-state convection in removing heat from a planetary interior, or lack thereof, plays a critical role. In Pluto's case, an ocean may exist between its surface ice shell and its rock (+metal + refractory organics) core (McKinnon et al., 2008). Specifically, if the deep ice is convecting, and the ocean contains *no* 'antifreezes' such as ammonia or methanol, then an ocean may or may not exist (i.e., be thermodynamically stable). If the deep ice is not convecting, but instead transports heat by conduction only, then an ocean must exist because the temperature gradient will be much larger than in the convecting case and intersect the ice melting curve for present-day heat flows (see McKinnon, 2006). Furthermore, if there is dissolved ammonia and/or methanol in Pluto's ocean, then it is extremely difficult (if not impossible) to freeze the ocean completely. In this case, and depending on the concentration of these species, Pluto's conductive ice shell could be ~200–300 km thick, overlying an ocean perhaps 100 km thick, albeit one cold and chemically noxious (Hussmann et al., 2006).

Charon in contrast is a modest-sized body and one that is not being tidally heated, so nominally, its interior should be rather cold today. Even if differentiated, it should not sustain an internal ocean (McKinnon et al., 2008; Schubert et al., 2010). The discovery of ammonia–water ice in Charon's spectrum (**Section 10.19.2.1**) has, however, prompted theoretical models that provide for survival of ammonia–water melt deep inside Charon up until the present day, for a narrow range of parameters (Cook et al., 2007; Desch et al., 2009).

Determining Pluto's (or Charon's) state of differentiation, much less the existence of oceans, will not be easy. *New Horizons* will not pass close enough to measure the degree-2 gravity field for either body (nor was such an original measurement requirement or mission objective), but shape determination from imaging offers the possibility of determining differentiation state. For this to be successful, however, Pluto or Charon must be in hydrostatic equilibrium. For all hydrostatic models of these two bodies, the differences between the lengths of the principal axes of their figures are predicted to be less than 1 km, and the differences between differentiated and undifferentiated models smaller still (McKinnon, 2013). Thus, discriminating between the two cases will be an observational challenge. Furthermore, either Pluto or Charon may choose not to cooperate, if they retain fossil tidal figures from earlier in their mutual tidal evolution, in the manner of Iapetus (e.g., Singer and McKinnon, 2011). These fossil figures would be larger (more biaxial in the case of Pluto and more triaxial for Charon), and thus much easier to measure, but would nevertheless provide important clues to the thermal and structural evolution of both bodies and Pluto in particular (Robouchon and Nimmo, 2011).

The comparisons with Triton that *New Horizons* should provide will be important not only in their own right but also for what they will imply about the geologic evolution of large KBOs generally. Triton's active geology and young crater retention age (Schenk and Zahnle, 2007; Stern and McKinnon, 2000) are often linked to the extreme tidal heating Triton almost certainly suffered after its capture by Neptune, but that capture event likely occurred billions of years ago (Nogueira et al., 2011; and see discussion in Schubert et al., 2010). From this perspective, then, Triton's recent activity would have to be powered instead by long-lived radiogenic heat (U, Th, or ^{40}K), acting in concert with a surface rich in very volatile (and hence geologically mobile) materials. And if so, Pluto should behave similarly today.

The specter of Triton's activity being driven by tidal heating has been recently resurrected, however. Although Triton's orbital eccentricity is essentially zero, because its orbit is highly inclined (157°) and retrograde, it has a small obliquity of order 0.5° (meaning its spin vector is slightly offset from its orbit normal). Jankowski et al. (1989) estimated that any heating due to solid-body tides raised by Neptune within Triton should be unimportant compared with radiogenic heating. Nimmo and Spencer (2014) proposed in contrast that tidal heating in a putative *ocean* within Triton supplies a substantial amount of heat. This is not a heat source available to Pluto or other large KBOs, so if this is the case, Pluto's level of geologic vigor will not match that of Triton today.

10.19.3 The Largest Trans-Neptunian Dwarf Planets and Dwarf Planet Candidates

There are at least a baker's dozen of large TNOs in the Pluto and Charon size class, including four presently recognized by the IAU as 'dwarf planets' (Eris, Makemake, Haumea, and Pluto of course) (Brown, 2008; Grundy et al., 2009; Sheppard et al., 2011). **Table 2** lists a 'top ten' KBOs in terms of size, as determined by a variety of methods, of which the stellar occultation method is deemed the most definitive. Other than Pluto, none of these bodies are either close enough to the Sun at present to raise, or are they large enough to sustain, a substantial atmosphere that could compromise measurement of a stellar occultation chord from the Earth. These relatively large bodies are all likely to have achieved hydrostatic equilibrium at low spherical harmonic degree, at least at some point in their histories, based on comparisons with icy satellites of similar size; thus, they are all probably dwarf planets (except for Charon, which is excused on a technicality). Ceres is included for comparison and to emphasize that the largest asteroid is an order of magnitude less massive than the largest KBOs.

Large KBOs are found in all of the dynamical groups or (sub)classes in the Kuiper Belt (Sheppard et al., 2011), although the great majority are found in the dynamically excited (or 'hot') resonant, scattered, and detached classes (Brown, 2008) (see **Section 10.19.4** for further discussion of these classes). In terms of the bodies in **Table 2**, Pluto–Charon and Orcus are both trapped in the 3:2 mean-motion resonance with Neptune (which makes Orcus a *plutino*).

Makemake, Haumea, and 2002 MS_4 are all part of the dynamically hot (high-inclination) component of the main, or classical, Kuiper Belt. Quaoar, at $i = 9°$, is often counted with this high-inclination group, but might be better thought of as an outlier of the dynamically 'cold' or low-inclination classical group, given its low eccentricity (0.04) and very red visible colors (a hallmark of the cold classical population) (Sheppard et al., 2011).

Eris, the most massive KBO known, is quite distant, but its orbit takes it close enough to Neptune at perihelion that it is generally viewed as a scattered disk object (i.e., Sheppard et al., 2011), although in the present epoch, it is no longer *scattering* and is formally classified as dynamically detached (Gladman et al., 2008); 2007 OR_{10}, with similar orbital elements, actually comes closer to the Sun than Eris but is protected from approaching Neptune by being in a mean-motion resonance with Neptune, albeit a weak, higher-order one (the 10:3).

Sedna is the most dynamically unusual TNO in **Table 2**, as its semimajor axis attests. It is dynamically *isolated*. With a perihelion of $\approx$75 AU, Sedna does not interact with Neptune or the other giant planets, yet its orbit is still too close to the Sun to be appreciably affected by passing stars or galactic tides (Brown, 2008; Gladman et al., 2008; and see Sheppard et al., 2011). The moniker 'detached' is an understatement.

10.19.3.1 Surface Compositions

The largest KBOs are also the brightest, generally, and thus can be spectroscopically interrogated. They are also generally higher-albedo bodies compared with the Kuiper Belt as a whole (Brown, 2008; Sheppard et al., 2011), a plausible consequence of geologic activity bringing ices and volatiles to their surfaces at some point (see McKinnon et al., 2008). Identified ices on a given body tend to be dominated spectrally by either H_2O or CH_4, whereas fainter KBOs tend to be spectrally featureless (e.g., Barucci et al., 2008a; Trujillo et al., 2011). Small band shifts of methane lines (a few to a few 10s of Å) are interpreted as a dilute solid-state solution of CH_4 in N_2 ice (as on Pluto and Triton), or vice versa (N_2 contamination of CH_4 ice), even if the 2.15 μm N_2 overtone feature cannot itself be detected (e.g., Tegler et al., 2008). Ethane (C_2H_6) ice is an expected photolytic and radiolytic alteration product of methane, and its detection is taken as corroboration that methane is indeed present on a given body.

The overall pattern of volatile identification can be understood, to first order, by considering both a KBO's size, and hence its ability to retain volatiles against escape, and its orbit, which determines surface temperature and surface temperature variation (Schaller and Brown, 2007b). Temperature determines which ices are likely to sublimate and raise a sensible atmosphere in vapor-pressure equilibrium, and insolation variation drives surface–atmosphere interactions that can cycle and refresh surfaces on seasonal timescales (e.g., Young, 2013). **Figure 4** illustrates how these factors interact. Methane ice is very easily seen spectroscopically and is detected on all the large, labeled KBOs in the figure to the right of the CH_4 stability line, with the

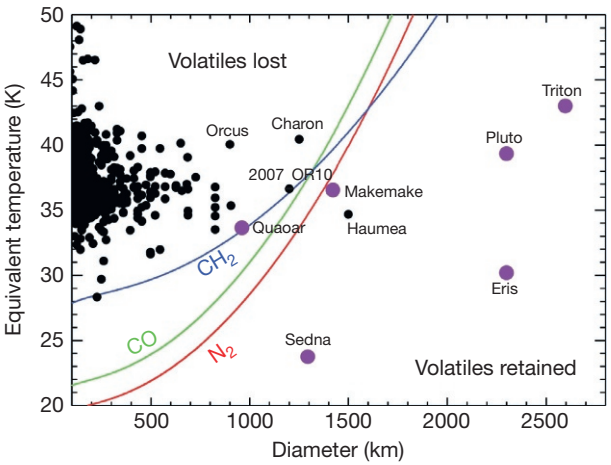


Figure 4 Volatile retention and loss in the Kuiper Belt. Objects to the left of the CH_4, CO, and N_2 lines are too small and too hot to retain any of those surface volatiles against Jeans escape over the age of the solar system, while objects to the right can retain those volatiles. All objects shown in purple have had CH_4 measured on their surfaces, although the detection of CH_4 on Sedna is less than conclusive (Barucci et al., 2010). In addition, Pluto, Triton, Eris, and Makemake have had N_2 or CO detected or inferred. No objects to the left of the lines have had any of these volatiles definitively detected. Methane is strongly suspected on 2007 OR_{10} (Brown et al., 2011). Note that diameters shown are not necessarily the same as (the more recent values) in **Table 2**, and equivalent temperature is that which provides the same escape flux as integration over an orbital period. Modified from Schaller EL and Brown ME (2007b) Volatile loss and retention on Kuiper Belt objects. *Astrophysical Journal Letters* 659: L61–L64; Brown ME, Burgasser AJ, and Fraser WC (2011) The surface composition of large Kuiper Belt object 2007 OR10. *Astrophysical Journal Letters* 738: L26 (4 pp.).

exception of Haumea (discussed in the succeeding text). Methane is a well-determined component of cometary ices (Mumma and Charnley, 2011), so its presence on the surfaces of KBOs that are more than large enough to have been internally active at some time (McKinnon et al., 2008) is not a surprise.

Nitrogen ice is perhaps more interesting. N_2 and CO are both 'supervolatile,' and whereas CO is a known component of cometary ices, molecular N_2 is not (Bockelée-Morvan et al., 2004) (N_2 is also notoriously difficult to detect; e.g., Cochran, 2002). Thus, the source of N_2 may be something else (and is commonly thought to be NH_3). If so, the presence of N_2 on Pluto and Triton (**Section 10.19.2.2**) and its inferred presence on Makemake and Eris (Tegler et al., 2008, 2010) may also be hallmark of large KBO thermochemistry. Sedna is large enough and cold enough for N_2 (or CO) to persist on its surface (cf. Barucci et al., 2005, 2010), but as Brown (2012) pointed out, Sedna is so cold that any nitrogen ice should be in the α phase and almost impossible to detect because of the extreme narrowness of its 2.15 μm absorption compared with that of β nitrogen. At its distance from the Sun, Sedna is a difficult target for spectroscopy (low signal/noise), and even the absorption features attributed to CH_4 are only marginally detected (Barucci et al., 2010).

Regarding the transitional bodies in **Figure 4** – Quaoar, 2007 OR_{10}, and Makemake – these bodies should have retained methane ice, but lost much of their allotment N_2 or CO. This is consistent with spectral information, especially considering that 2007 OR_{10} and Quaoar are likely somewhat larger than plotted in the figure (Braga-Ribas et al., 2013; Santos-Sanz et al., 2012). All three bodies are spectrally very red (a plausible signature of methane irradiation), and Quaoar and Makemake have methane *and* ethane on their surfaces (Brown et al., 2007a; Dalle Ore et al., 2009; Schaller and Brown, 2007a). Quaoar and 2007 OR_{10} also exhibit the signature of H_2O ice (Brown et al., 2011; Jewitt and Luu, 2004), which indicates that these two bodies have reached an advanced stage of methane loss. Makemake, in contrast, has a very methane-rich surface (Brown et al., 2007a; Licandro et al., 2006) and still retains a modest amount of N_2 (or CO), based on very small band shifts in its methane absorptions (Tegler et al., 2008). These small shifts imply either frost patches of N_2-rich ice or, perhaps more likely, small amounts of N_2 dissolved in CH_4-ice crystals as a contaminant. This is consistent with Makemake's position on the figure.

Charon and Haumea sit on opposite sides of the stability lines in **Figure 4**. Both are spectrally dominated by water ice absorptions (Barucci et al., 2008a; Brown, 2008). Haumea could, conceivably, stably host N_2 or CO ices on its surface. This remarkable KBO is orbited by two known satellites and itself orbits adjacent to a dynamical family of fragments, all of which bear clear, deep spectral signatures of crystalline water ice (Brown et al., 2007b). Haumea is also rather dense and rotating so rapidly (with a 3.9 h period) that it has apparently distorted into a Jacobi-figure-like triaxial body. This state of affairs is most elegantly explained by a massive collision in which a differentiated proto-Haumea lost much of its icy mantle (Brown et al., 2007b) or in which both impactor and target lost substantial ice (Leinhardt et al., 2010) (see also Schlichting and Sari, 2009; Ortiz et al., 2012; Ćuk et al., 2013). Presumably, any volatile-ice-rich crust was lost along with ejected mantle, leaving mostly water ice behind.

Charon, of course, is also likely the product of a massive collision, so its water ice surface may be similarly explained. But Charon is a smaller and warmer body than Haumea, and CH_4, N_2, and CO would not be stable on its surface over the age of the solar system regardless, according to the simple model shown. On the other hand, Charon is spectrally neutral in the visible (Buie et al., 2010b); this may imply that Charon's surface ices have always lacked CH_4 or other light hydrocarbons such as methanol. Otherwise, UV and radiation processing would surely have reddened some of the hydrocarbons, permanently reddening Charon's surface.

Finally, Orcus, in the upper left of **Figure 4**, shares with Charon a flat, spectrally neutral visible spectrum and well-defined water ice absorptions at 1.6 and 2.0 μm (Brown, 2008). There is also a smaller absorption at 2.2 μm that is reminiscent of Charon's spectrum at that wavelength, but the signal-to-noise ratio is such that a clear identification (e.g., of ammonia) cannot yet be made (Barucci et al., 2008b; Carry et al., 2011; Delsanti et al., 2010). Overall, the utility of **Figure 4** for understanding volatile-ice compositional patterns is remarkable and suggests common cosmochemical processes are at work.

As an aside, one of these cosmochemical processes must be the preservation of crystalline water ice against radiation-induced amorphotization. The 1.65 μm absorption feature specific to *crystalline* water ice is ubiquitous among KBOs and icy satellites where H_2O ice is seen. A number of authors, when first detecting this band, have implicated recent icy or cryovolcanism, regardless of the size or context of the body in question. That midsize and larger KBOs possess crystalline water ice should not be surprising: these bodies are all of the size that impact and/or long-term radiogenic heating would have caused any amorphous water ice originally accreted (and which is, after all, metastable) to crystallize (e.g., McKinnon, 2002; McKinnon et al., 2008). Water ice exposed at the surface to the space environment and thus to ionizing radiation may, however, be expected to reamorphotize (Cook et al., 2007), yet this is not seen. Of course, ionizing UV radiation does not penetrate deeply, and cosmic rays may penetrate too deeply to compete with impact-driven regolith gardening. Clearly, our understanding is incomplete.

10.19.3.2 Densities and Interiors

As important as surface compositions are, bulk densities are at as least as important for constraining formation, origin, and composition. Taken together, they provide deeper insight into solar nebula chemistry and planetary formation processes (see, e.g., **Chapter 10.02** in this volume). Although it has been historically difficult to determine bulk densities accurately, recent years have seen the discovery of satellites for a number of large and midsize KBOs (some of the latter are sometimes listed as Centaurs, but here, we follow the definition of Gladman et al. (2008) that Centaur semimajor axes must be less than that of Neptune). With the discovery of satellites, or binarity, Kepler's 3rd law gives the system mass and, with size estimates from occultations or radiometry, densities.

Figure 5 illustrates densities of TNOs for which constraining mass and radius estimates are available. Typhon, Ceto, Sila, and 1998 SM_{165} are all binaries, while 1999 TC_{36} is a triple system, which complicates radiometric interpretations of diameter.

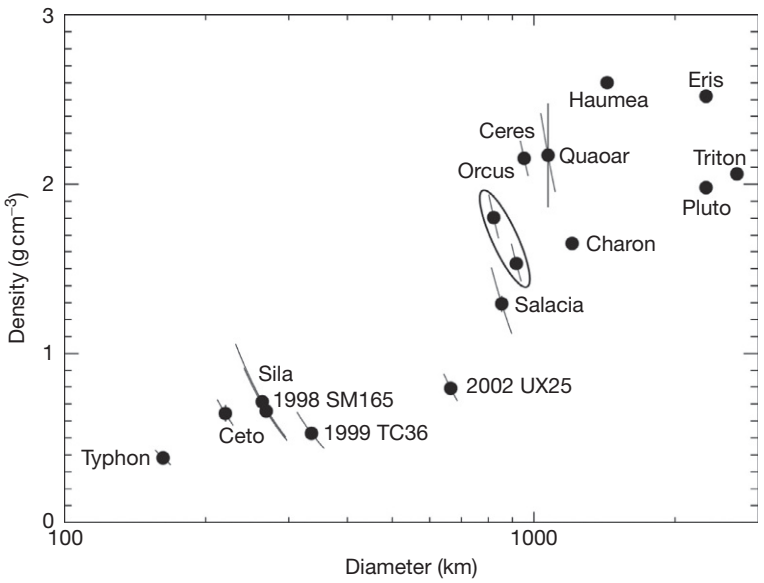


Figure 5 Densities of large and midsize Kuiper Belt objects (KBOs) and Ceres. Uncertainties in diameter generally dominate uncertainties in mass. Diameters from radiometry (midsize KBOs) are judged less reliable than those from stellar occultations (most of the large KBOs and Ceres). The correlation of density with diameter cannot simply be a function of decreasing porosity with increasing body size, but implies a trend of decreasing iciness as size increases (see text). Note that, as with **Figure 4**, diameters shown are not necessarily the same as in **Table 2**. Modified from Brown ME (2013) The density of mid-sized Kuiper belt object 2002 UX25 and the formation of the dwarf planets. *The Astrophysical Journal Letters* 778: L34 (5 pp.).

The apparent correlation of KBO size with density has been noted by a number of authors. As discussed by McKinnon et al. (2008), one simple explanation that has been offered is that all KBOs have a solid or grain density near 2 g cm^{-3} – a value independently supported by cosmochemical arguments (e.g., Wong et al., 2008) – but that smaller KBOs are substantially porous. There are, however, limits to the maximum porosity that can be reasonably, frictionally supported in any midsize 'rubble-pile' KBO, probably about 40% and decreasing with increasing pressure (Durham et al., 2005). The inference, then, is that all the low-density bodies in **Figure 5** are substantially icier than, for example, Pluto, Triton, or Eris.

What alternative process could then lead to the size–density correlation in **Figure 5**? One clue is offered by Haumea. It appears to be the densest known KBO at present and, as discussed previously, is thought to have lost up to half its icy mantle in a massive impact (Brown et al., 2007b). Haumea's satellites and family members presumably could be water ice-rich (consistent with their spectra), porous, and of low density. If this phenomenon can be generalized, then large KBOs, if differentiated, would tend to evolve to be more rock-rich, while any ejected icy fragments would tend to make the smaller and midsize KBO population more ice-rich as a whole. Brown (2012) outlined a simple, heuristic model of this process, but concluded that the ice/rock fractionation process was too inefficient to explain the observations.

The size–density correlation in **Figure 5**, if real, is clearly an important one and offers an important clue as to how the Kuiper Belt formed. For example, if large KBOs formed by the traditional hierarchy of two-body, accretionary collisions (coagulation) (Kenyon and Bromley, 2012), and KBO-forming planetesimals all started suitably rock-rich, then as Brown (2012) determined, it may be difficult to create enough ice-rich, smaller KBOs to much affect the composition of the small and midsize population as a whole. If, however, KBOs are all 'born big,' that is, formed from the gravitational collapse of large aggregates of very small 'pebbles' concentrated by streaming instabilities in the solar nebula (Johansen et al., 2012; cf. Morbidelli et al., 2009), then most smaller KBOs should be fragments created by impacts of and on these larger KBOs. This second generation of smaller KBOs could then be quite icy by their very nature, but only if the original, large KBOs had differentiated early on.

The issue of the compositions of small and midsize KBOs is connected to the issue of the compositions of comets (A'Hearn, 2011). It is certainly not obvious that comets are collisional fragments as opposed to being primordial planetesimals (Mumma and Charnley, 2011), although estimates of cometary densities are certainly compatible with the low densities in **Figure 5**. (And it should be noted that the internal stresses in 1–10 km diameter comet nuclei are so low that almost any porosity can be maintained by internal strength.) Nor is the robustness of the correlation in **Figure 5** established. Densities based on direct imaging or radiometric modeling have tended to be replaced over time by more accurate stellar occultation values. For example, Quaoar is no longer viewed as rocky as estimated initially (Fraser and Brown, 2010). Also, Phoebe, the 210 km diameter irregular satellite of Saturn, is widely thought to have been captured from the Kuiper Belt, and its mean density is well known at 1.64 g cm^{-3} (see Wong et al., 2008). More accurate measurements of midsize KBO densities are needed.

10.19.4 The Kuiper Belt

The small body populations of the solar system are the archaeological or 'fossilized' remnants of what were once vast

and innumerable planetesimal swarms. The extant trans-Neptunian population, viewed broadly, stretches from the Jupiter-family comets, through the Centaurs, across the 'classical' KBOs outside Neptune's orbit, to scattering and detached deep space objects well beyond (Gladman et al., 2008; **Figure 6**). As of this writing, over 1650 TNOs are known, but the actual population is much larger, perhaps 200 000 bodies of 100 km diameter or more and with a combined mass of 0.05–0.10 Earth masses (Fuentes and Holman, 2008; Gladman et al., 2012; Kaib et al., 2011; Petit et al., 2011). The dynamical complexity of the entire Kuiper Belt is evident in **Figure 6**, and understanding its details has opened a window to understanding the origin and evolution of the early solar system that never existed before.

As alluded to earlier, the Kuiper Belt population can be divided into two broad populations, dynamically excited (in terms of eccentricity e or inclination i), or 'hot,' and much less dynamically excited (low e and i), or 'cold.' Further distinctions are physically meaningful (**Figure 6**). Following Petit et al. (2011), resonant objects are those currently in a mean-motion resonance with Neptune, and scattering objects are those which over 10 My integrations experience encounters with Neptune resulting in a semimajor axis a deviation of more than 1.5 AU. Classical bodies can be further subdivided into the inner classical objects (a interior to the 3:2 mean-motion resonance), main classical objects (a between the 3:2 and 2:1 mean-motion resonances), and outer classical objects (a exterior to the 2:1 mean-motion resonance). Classical objects can also be categorized as dynamically hot or cold, but all the cold classicals ($i < 5°$ or so) are in the main group.

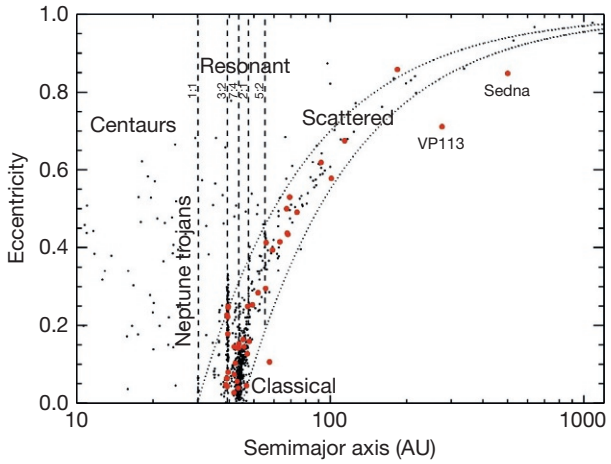


Figure 6 Semimajor axis versus eccentricity of multiopposition trans-Neptunian objects (updated from Sheppard et al., 2011; S. Sheppard, personal communication). Large circles represent TNOs with absolute magnitude $H \leq 4.5$ magnitude (including those in **Table 2**). Several distinct dynamical KBO populations are shown. Vertical dashed lines indicate the main mean-motion resonances with Neptune as well as the Neptune Trojans in the 1:1 resonance. In this figure, 'scattered disk' objects have perihelia between 30 and 45 AU as shown between the dashed lines, but this does not have the same meaning as scattering objects, as described in the text. 'Classical' objects are in the lower center portion of the figure and include the high- and low-inclination populations. An edge near 48 AU can clearly be seen for low-eccentricity objects. Centaurs are on unstable orbits between the giant planets. Sedna and 2012 VP_{113} stand out as being significantly below the perihelion line shown at 45 AU and are dynamically decoupled from Neptune, as discussed in **Section 10.19.2**. Some authors (e.g., Trujillo and Sheppard, 2014) regard these two bodies as inliers of a much more massive, inner Oort cloud population.

Detached objects, such as Sedna and 2012 VP_{113} (Trujillo and Sheppard, 2014), have a beyond the 2:1 mean-motion resonance as well as high eccentricities, though the distinction between detached objects and outer classicals is somewhat arbitrary (an eccentricity cut) (Petit et al., 2011). Finally, those 'KBOs' on dynamically unstable orbits between the giant planets (and with $a < a_{\text{Neptune}}$) are classified as Centaurs. Of the 1650 known TNOs in the preceding text, about 385 are Centaurs or scattered disk objects, and 9 are Neptune Trojans (www.minorplanetcenter.net). And it is worth remembering that, with the exception of the detached objects, KBO orbits generally *overlap* in physical space but not in orbital parameter space (**Figure 6**).

10.19.4.1 Nice Models

The family of dynamical models known as 'Nice' (after the city of their birth) has taken a broad variety of ideas in planetary formation and attempted to create a comprehensive working model of what actually might have happened during the solar system's earliest epochs (e.g., Morbidelli et al., 2008). It takes as a point of departure that the solar system was once much more compact, in the sense that the giant planets were all located between ~5 and 15 AU, and there was a remnant planetesimal disk that stretched out to ~30–35 AU (Tsiganis et al., 2005). Slow evolution (migration) of the giant planets' orbits (see Hahn and Malhotra, 1999; Malhotra, 1995) eventually leads to resonance crossing by Jupiter and Saturn, which creates systemic orbital instability. This instability expels the ice giants into the planetesimal disk, and after which, all four giant planet orbits settle down by means of dynamical friction (exchange of orbital energy) with the planetesimals (**Figure 7**).

In the original quantitative 'proof-of-concept' calculations, the secular architecture (a, e, and i) of the giant planets was correctly reproduced (Tsiganis et al., 2005). The model had the additional attribute of providing a mechanism for creating a 'Late Heavy Bombardment' in the inner solar system at ~3.9 Ga, provided the instability could be delayed (and the planetesimal disk survive) for some 600 My (Gomes et al., 2005). This does not mean that matters proceeded precisely as outlined in **Figure 7** (see, e.g., Batygin and Brown, 2010), but the Nice model and its extensions (e.g., Brasser et al., 2009; Levison et al., 2011; Walsh et al., 2011) have proved remarkably adept at explaining various features of solar system architecture. Most relevant to the present discussion is the structure of the Kuiper Belt.

Figure 8 shows the outcome of a numerical simulation in which the Kuiper Belt is filled by chaotically scattered planetesimals in the original formulation of the Nice model ('Nice I') (Levison et al., 2008). The Kuiper Belt region in this model is initially *empty*. The correspondence between the observed and simulated KBOs in **Figure 8** is qualitatively and even quantitatively rather good, once observational biases are taken into account (i.e., distant, as well as high-e or high-i, objects are less likely to be detected in a given survey). Specifically, the

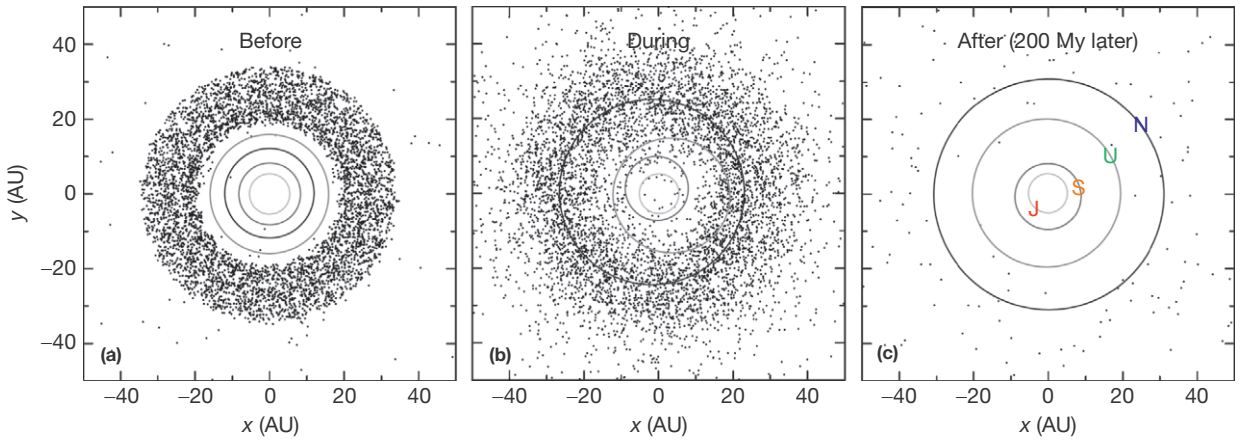


Figure 7 Evolution of a compact, primordial giant planet architecture and surrounding planetesimal disk, in the original demonstration of orbital migration and cataclysmic instability now known as the 'Nice model.' The mass (~30 M_{Earth}) and outer truncation of the planetesimal disk near 35 AU are critical elements in getting the correct, final giant planet orbits. The scattered planetesimals represent the 'Kuiper belt.' Modified from Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.

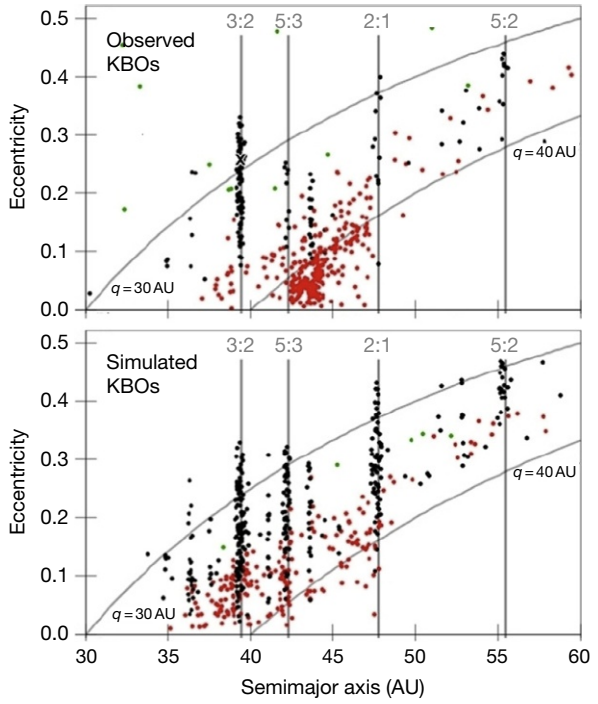


Figure 8 Eccentricity versus semimajor axis distributions for multiopposition KBOs (top) and the results of a Nice model installment simulation (bottom) (modified from Levison HF, Morbidelli A, VanLaerhoven C, Gomes R, and Tsiganis K (2008) Origin of the structure of the Kuiper belt during a dynamical instability in the orbits of Uranus and Neptune. *Icarus* 196: 258–273). Black dots represent resonant objects (with mean-motion resonances indicated), green dots are scattering objects, and red dots are stable, nonresonant objects; X marks Pluto. The simulation is a reasonable facsimile of the observed distribution, once observational biases are taken into account. The simulation displays a deficit of nearly circular objects in the classical Kuiper Belt, however, but does somewhat better with reproducing the spread in cold classical inclinations (not shown). Note the change in semimajor axis scale with respect to **Figure 6**.

mean-motion resonances are populated as well as the 'scattered disk' and hot classical regions. The overall structure in **Figure 8** is mostly due to Neptune, as it completes its outward migration and its eccentricity damps from larger values.

An important aspect of the Kuiper Belt that the simulation in **Figure 8** obviously does not reproduce, as Levison et al. (2008) discussed, is the cold classical population, especially the low-*e* populated region between the 5:3 resonance and 44 AU (**Figure 8**, top). One explanation is that this region contains survivors of a primordial planetesimal disk at this location. This is in some ways an attractive explanation, as the cold classicals are known to be physically smaller (lack large bodies, see next section) and to be spectrally redder than the hot populations (e.g., Doressoundiram et al., 2008). A much larger fraction of the cold classicals are wide binaries (10s of percent), compared with the hot classicals (Noll et al., 2008a,b). In other words, the cold classicals appear to be a physically *and* dynamically distinct population.

Analytic and numerical models have shown that Nice-like instabilities can be compatible with a preserved and only modestly dynamically excited cold classical population, provided Neptune's dynamical behavior conforms to certain strictures (Batygin et al., 2011; Dawson and Murray-Clay, 2012). On the other hand, too much mass in an extended planetesimal disk (out to ~48 AU) would mean that Neptune's outward migration would not halt near 30 AU, so other explanations for the cold classical 'kernel' may yet emerge (Morbidelli et al., 2014; Petit et al., 2011).

The prospect that the cold classical population may represent a primordial, *in situ* population is nonetheless exciting, for two reasons. First, as Batygin et al. (2011) pointed out, the cold classical population would be a "unique laboratory for the study of [disk] surface processes as well as the chemistry of the primordial solar nebula." Second, the *New Horizons* mission, in its extended post-Pluto phase, could fly by a cold classical KBO in the 25–50 km size range. As of this writing, a suitable KBO target has not been found, despite several years of

ground-based searches (Buie et al., 2012a; Parker et al., 2013). Suitable in this case means a KBO that can be reached using the thruster fuel still on board the spacecraft. This search will be renewed in the summer of 2014.

10.19.4.2 Kuiper Belt Population Characteristics

The size distribution of the Kuiper Belt and its various populations are fundamental properties that reflect the accretional and collisional processes that created and shaped these bodies. A great deal of effort over the years has gone into estimating their size-frequency distributions (SFDs), as summarized in Petit et al. (2008). The difficulty lies in converting measurements of apparent luminosities, of distant objects subject to numerous observational biases and incompletenesses, into robust estimates of size. Recent efforts, however, appear to be converging in several important regards. The Canada-France Ecliptic Plane Survey (CFEPS) reports KBO detections from an exceptionally well-characterized survey from a single telescope (Petit et al., 2011). They find that the dynamically hot classicals follow a cumulative magnitude slope $\alpha_{\rm hot} \approx 0.8$, whereas the cold classicals are steeper with an $\alpha_{\rm cold} \approx 1.2$. These slopes can be converted into a differential power-law index of the form $dN/dR \propto R^{-q^*}$, where $q^* = 5\alpha + 1$, N is number of objects, and R is object radius (assuming uniform albedos). Both of these differential size distributions are steep when compared to the collisional slope of Dohnanyi (1969), $q^* = -3.5$ (for size-independent strength). Similar slopes for the cold and hot populations are reported by Fraser et al. (2014) from their multisurvey derivation of KBO absolute magnitudes.

In both Petit et al. (2011) and Fraser et al. (2014), the slopes earlier refer to the bright or large size end of the KBO populations in question. The steep slopes are considered to be a fossilized signature of the accretion of larger bodies from smaller ones (Kenyon and Bromley, 2012; Kenyon et al., 2008). Fraser et al. (2014) further reported that the magnitude distribution of fainter KBOs follows a rather shallow $\alpha_{\rm faint} \sim 0.2$–$0.4$ ($q^* \sim 2$–3), with a break diameter of ~110–140 km. The differences between the hot and cold populations and the inference for a break or knee in the SFD go back to the deep HST survey results of Bernstein et al. (2004), but there now appears to be agreement on this aspect of the population structure.

Analysis of deep *Subaru* imaging (Fraser and Kavelaars, 2009; Fuentes and Holman, 2008; Fuentes et al., 2009) also supports a knee or double power law for the TNO SFD. Shankman et al. (2013) further allowed (based on CFEPS data for the scattering population) that, rather than a simple knee in the SFD, there may be deficit or 'divot' of bodies below a break diameter of ~100 km and, at still smaller sizes, a steepening of the distribution (to perhaps $\alpha_{\rm faint} \sim 0.5$). At the very faint end (<10 km diameter), there are very few direct observational constraints, and some rather complex SFDs have been proposed (Schlichting et al., 2013).

An important aspect of all this work is that similarly steep population indices appear to characterize all the hot populations, classical, scattered, and resonant, at the bright ends of their distributions (Fraser et al., 2014; Gladman et al., 2012; Petit et al., 2011; Schwamb et al., 2014). Moreover, Fraser et al. (2014) found that the known, broken power-law SFD of the Jovian Trojan asteroids is a statistical match with their hot population model (i.e., the hypothesis that the hot population and the Trojans were drawn from the same parent population cannot be rejected). Capture of Trojan populations is a prediction of the Nice model (Morbidelli et al., 2005; Nesvorný et al., 2013), so these populations should be similar (barring strong, separate collisional evolution). A further implication is that the cold classicals are truly a different, and possibly *in situ*, population, as discussed earlier.

It remains to be seen, however, how the different color (spectral slope) characteristics of the resonant KBO populations can be explained (Sheppard, 2012). Naively, they should all share the same 'hot population' color characteristics, but for example, objects in the 5:3 and 7:4 resonances are mostly of the 'ultrared' variety, which is characteristic of the low-inclination, cold classicals. These two resonances are, however, anchored (in terms of semimajor axis) in the core of the cold classical population (**Figure 8**, top). Indeed, they may actually have been populated from the cold classicals, as Neptune completed its migration (Batygin et al., 2011; B. Gladman, personal communication), as opposed to being populated by bodies transported from closer to the Sun, as in Levison et al. (2008).

There is clearly much to learn about how the resonant populations of Kuiper Belt were created and what that implies for the global evolution of the solar system. For example, based on CFEPS detections, Gladman et al. (2012) estimated that resonances beyond the 2:1 are more populated than Nice model simulations have shown to date. Deeper surveys will be necessary to put our understanding of these more distant, resonant populations on a firmer footing.

10.19.5 Final Remarks

The revelation of the Kuiper Belt is arguably the most important advance in solar system science in the last 30 years. Its only rival would be the discovery and characterization of planets around other stars. Yet, taken together, these represent a paradigm shift in terms of the stability and nominally uniformitarian evolution of our solar system. Most researchers now recognize that the solar system was likely much different (compact) and far more dynamic (violent) in its first 600 million years. The Kuiper Belt contains the richest record of this bygone era. Dwarf planets such as Pluto likely accreted well inside their present orbital positions and have much to teach us about the chemical and physical conditions in the earlier trans-Neptunian disk. Our understanding of the Pluto–Charon system should improve dramatically in 2015, and in 2016, because the data downlink from the July 2015 *New Horizons* encounter will take over a year. Beyond that, the coming generation of ground-based and space-based telescopes should probe trans-Neptunian space ever deeper and more completely. The full story of Pluto and Kuiper Belt is yet unwritten.

Acknowledgments

Support for this introduction and review was provided by NASA and the *New Horizons* project. I thank Brett Gladman, Will Grundy, Scott Sheppard, and Tilman Spohn for their comments. Deep searches by *HST* in summer 2014 have located one, and possibly two, post-Pluto KBO tragets for *New Horizons*.

References

A'Hearn MJ (2011) Comets as building blocks. *Annual Review of Astronomy and Astrophysics* 49: 281–299.

Atreya SK, Lorenz RD, and Waite JH (2009) Volatile origin and cycles: Nitrogen and methane. In: Brown RH, Lebreton J-P, and Waite JH (eds.) *Titan from Cassini-Huygens*, pp. 77–99. Dordrecht, Heidelberg, London, New York: Springer. http://dx.doi.org/10.1007/978-1-4020-9215-2_7.

Barucci MA, Brown ME, Emery JP, and Merlin F (2008a) Composition and surface properties of transneptunian objects and Centaurs. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 143–160. Tucson, AZ: University of Arizona Press.

Barucci MA, Cruikshank DP, Dotto E, et al. (2005) Is Sedna another Triton? *Astronomy & Astrophysics* 439: L1–L4.

Barucci MA, Dalle Ore CM, Alvarez-Candal A, et al. (2010) (90377) Sedna: Investigation of surface compositional variation. *Astronomical Journal* 140: 2095–2100.

Barucci MA, Merlin F, Guilbert A, et al. (2008b) Surface composition and temperature of the TNO Orcus. *Astronomy & Astrophysics* 479: L13–L16.

Batygin K and Brown ME (2010) Early dynamical evolution of the solar system: Pinning down the initial conditions of the Nice model. *Astrophysical Journal* 716: 1323–1331.

Batygin K, Brown ME, and Fraser WC (2011) Retention of a primordial cold classical Kuiper belt in an instability-driven model of solar system formation. *Astrophysical Journal* 738: 13 (8 pp.).

Bernstein GM, Trilling DE, Allen RL, Brown ME, Holman M, and Malhotra R (2004) The size distribution of trans-Neptunian bodies. *Astronomical Journal* 128: 1364–1390.

Binzel RP and Hubbard WB (1997) Mutual events and stellar occultations. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 85–102. Tucson, AZ: University of Arizona Press.

Bockelée-Morvan D, Crovisier J, Mumma MJ, and Weaver HA (2004) The composition of cometary volatiles. In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 391–423. Tucson, AZ: University of Arizona Press.

Bosh AS, Person MJ, Levine SE, et al. (2013) The state of Pluto's atmosphere in 2012–2013. *AAS/DPS Meeting 45*, 404.01.

Braga-Ribas F, Sicardy B, Ortiz JL, et al. (2013) The size, shape, albedo, density, and atmospheric limit of transneptunian object (50000) Quaoar from multi-chord stellar occultations. *Astrophysical Journal* 773: 26 (13 pp.).

Brasser R, Morbidelli A, Gomes R, Tsiganis K, and Levison HF (2009) Constructing the secular architecture of the Solar System II: The terrestrial planets. *Astronomy & Astrophysics* 507: 1053–1065.

Bromley SJ and Kenyon BC (2014) The formation of Pluto's low-mass satellites. *Astronomical Journal* 147: 8 (18 pp.).

Brown ME (2008) The largest Kuiper belt objects. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 335–344. Tucson, AZ: University of Arizona Press.

Brown ME (2012) The compositions of Kuiper belt objects. *Annual Review of Earth and Planetary Sciences* 40: 467–494.

Brown ME (2013) The density of mid-sized Kuiper belt object 2002 UX25 and the formation of the dwarf planets. *The Astrophysical Journal Letters* 778: L34 (5 pp.).

Brown ME, Barkume KM, Blake GA, et al. (2007a) Methane and ethane on the bright Kuiper belt object 2005 FY9. *Astronomical Journal* 133: 284–289.

Brown ME, Barkume KM, Ragozzine D, and Schaller EL (2007b) A collisional family of icy objects in the Kuiper belt. *Nature* 446: 294–296.

Brown ME, Burgasser AJ, and Fraser WC (2011) The surface composition of large Kuiper belt object 2007 OR10. *Astrophysical Journal Letters* 738: L26 (4 pp.).

Brown ME and Calvin WM (1999) Evidence for crystalline water and ammonia ices on Pluto's satellite Charon. *Science* 287: 107–109.

Brozović M, Showalter M, Jacobson RA, and Buie M (2014) The orbits and masses of satellites of Pluto. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2014.03.015.

Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2010a) Pluto and Charon with the *Hubble Space Telescope*. I. Monitoring global change and improved surface properties from light curves. *Astronomical Journal* 139: 1117–1127.

Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2010b) Pluto and Charon with the *Hubble Space Telescope*. II. Resolving changes on Pluto's surface and a map for Charon. *Astronomical Journal* 139: 1128–1143.

Buie MW, Spencer JR, Parker AW, et al. (2012a) Searching for KBO flyby targets for the New Horizons mission. *Asteroids, Comets, Meteors 2012b*, abstract #6430.

Buie MW, Tholen DJ, and Grundy WM (2012a) The orbit of Charon is circular. *Astronomical Journal* 144: 15 (19 pp.).

Buie MW, Tholen DJ, and Wasserman LH (1997) Separate light curves of Pluto and Charon. *Icarus* 125: 233–244.

Canup RM (2005) A giant impact origin of Pluto–Charon. *Science* 307: 546–550.

Canup RM (2011) On a giant impact origin of Charon, Nix, and Hydra. *Astronomical Journal* 141: 35 (9 pp.).

Carry B, Hestroffer D, and DeMeo FE (2011) Integral-field spectroscopy of (90482) Orcus-Vanth. *Astronomy & Astrophysics* 534: A115 (9 pp.).

Cochran AL (2002) A search for N_2^+ in spectra of comet C/2002 C1 (Ikeya-Zhang). *The Astrophysical Journal Letters* 576: L165–L168.

Cook JC, Desch SJ, Roush TL, Trujillo CA, and Geballe TR (2007) Near infrared spectroscopy of Charon: Possible evidence for cryovolcanism on Kuiper belt objects. *Astrophysical Journal* 663: 1406–1419.

Cruikshank D, Peale SJ, and Harris AW (1997) The surfaces of Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 221–267. Tucson, AZ: University of Arizona Press.

Ćuk M, Ragozzine D, and Nesvorný D (2013) On the dynamics and origin of Haumea's moons. *Astronomical Journal* 146: 89 (13pp.).

Dalle Ore CM, Barucci MA, Emery JP, et al. (2009) Composition of KBO (50000) Quaoar. *Astronomy & Astrophysics* 501: 349–357.

Dawson RI and Murray-Clay RA (2012) Neptune's wild days: Constraints from the eccentricity distribution of the classical Kuiper belt. *Astrophysical Journal* 750: 43 (29 pp.).

de Bergh C, Schmitt B, Moroz LV, Quirico E, and Cruikshank DP (2008) Laboratory data on ices, refractory carbonaceous materials, and minerals relevant to transneptunian objects and Centaurs. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 483–506. Tucson, AZ: University of Arizona Press.

Delsanti A, Merlin F, Guilbert-Lepoutre A, Bauer J, Yang B, and Meech KJ (2010) Methane, ammonia, and their irradiation products at the surface of an intermediate-size KBO? A portrait of Plutino (90482) Orcus. *Astronomy & Astrophysics* 520: A40 (15 pp.).

DeMeo FE, Dumas C, de Bergh C, et al. (2010) A search for ethane on Pluto and Triton. *Icarus* 208: 412–424.

Desch SJ, Cook JC, Doggett TC, and Porter SB (2009) Thermal evolution of Kuiper belt objects, with implications for cryovolcanism. *Icarus* 202: 694–714.

Dobrovolskis AR, Peale SJ, and Harris AW (1997) Dynamics of the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 159–189. Tucson, AZ: University of Arizona Press.

Dohnanyi JS (1969) Collisional model of asteroids and their debris. *Journal of Geophysical Research* 74: 2531–2554.

Doressoundiram A, Boehnhardt H, Tegler SC, and Trujillo C (2008) Color properties and trends of the transneptunian objects. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 91–104. Tucson, AZ: University of Arizona Press.

Douté S, Schmitt B, Quirico E, et al. (1999) Evidence for methane segregation at the surface of Pluto. *Icarus* 142: 421–444.

Dumas C, Terrile RJ, Brown RH, Schneider G, and Smith BA (2001) *Hubble Space Telescope* NICMOS spectroscopy of Charon's leading and trailing hemispheres. *Astronomical Journal* 121: 1163–1170.

Durham WB, McKinnon WB, and Stern LA (2005) Cold compaction of water ice. *Geophysical Research Letters* 32: L18202.

Elliot JL, Ates A, Babcock BA, et al. (2003) The recent expansion of Pluto's atmosphere. *Nature* 424: 165–168.

Elliot JL, Dunham EW, Bosh AS, et al. (1989) Pluto's atmosphere. *Icarus* 77: 148–170.

Elliot JL, Kern SD, Clancy KB, et al. (2005) The Deep Ecliptic Survey: A search for Kuiper belt objects and Centaurs. II. Dynamical classification, the Kuiper belt plane, and the core population. *Astronomical Journal* 129: 1117–1162.

Elliot JL, Person MJ, Gulbis AAS, et al. (2007) Changes in Pluto's atmosphere: 1988–2006. *Astronomical Journal* 134: 1–13.

Emery JP, Dalle Ore CM, Cruikshank DP, Fernández YR, Trilling DE, and Stansberry JA (2007) Ices on (90377) Sedna: Confirmation and compositional constraints. *Astronomy & Astrophysics* 466: 395–398.

Fornasier E, Lellouch E, Müller T, et al. (2013) "TNOs are Cool": A survey of the trans-Neptunian region VIII. Combined Herschel PACS and SPIRE observations of 9 bright targets at 70–500 μm. *Astronomy & Astrophysics* 555: A15 (22 pp.).

Fraser WC and Brown ME (2010) Quaoar: A rock in the Kuiper belt. *Astrophysical Journal* 714: 1547–1550.

Fraser WC and Kavelaars JJ (2009) The size distribution of Kuiper belt objects for $D \gtrsim 10$ *km*. *Astronomical Journal* 137: 72–82.

Fraser WC, Brown ME, Morbidelli A, Parker A, and Batygin K (2014) The absolute magnitude distribution of Kuiper belt objects. *Astrophysical Journal* 782: 100 (14 pp.).

Fuentes CI, George MR, and Holman MJ (2009) A Subaru pencil-beam search for $m_r \sim 27$ trans-Neptunian bodies. *Astrophysical Journal* 696: 91–95.

Fuentes CI and Holman MJ (2008) A Subaru archival search for faint trans-Neptunian objects. *Astronomical Journal* 136: 83–97.

Gladman B, Lawler SM, Petit J-M, et al. (2012) The resonant trans-Neptunian populations. *Astronomical Journal* 144: 23–47.

Gladman B, Marsden BG, and VanLaerhoven C (2008) Nomenclature in the outer solar system. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 43–57. Tucson, AZ: University of Arizona Press.

Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* 435: 466–469.

Grundy WM and Buie MW (2001) Distribution and evolution of CH_4, N_2, and CO ices on Pluto's surface: 1995 to 1998. *Icarus* 153: 248–263.

Grundy WM, McKinnon WB, Ammannito E, et al. (2009) *SBAG Community White Paper for the Planetary Science Decadal Survey*. Washington, DC: National Academy of Sciences.

Grundy WM, Olkin CB, Young LA, Buie MW, and Young EF (2013) Near-infrared spectral monitoring of Pluto's ices: Spatial distribution and secular evolution. *Icarus* 223: 710–721.

Grundy WM, Olkin CB, Young LA, and Holler BJ (2014) Near-infrared spectral monitoring of Pluto's ices II: Recent decline of CO and N_2 ice absorptions. *Icarus* 233: 220–224.

Grundy WM, Young LA, Stansberry JA, Buie MW, Olkin CB, and Young EF (2010) Near-infrared spectral monitoring of Triton with IRTF/SpeX II: Spatial distribution and evolution of ices. *Icarus* 205: 594–604.

Hahn JM and Malhotra R (1999) Orbital evolution of planets embedded in a planetesimal disk. *Astronomical Journal* 117: 3041–3053.

Hudson RL, Palumbo ME, Strazulla G, Moore MH, Cooper GF, and Sturner SJ (2008) Laboratory studies of the chemistry of transneptunian object surface materials. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 507–523. Tucson, AZ: University of Arizona Press.

Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-neptunian objects. *Icarus* 185: 258–273.

Jankowski DG, Chyba CF, and Nicholson PD (1989) On the obliquity and tidal heating of Triton. *Icarus* 80: 211–219.

Jewitt DC and Luu J (2004) Crystalline water ice on the Kuiper belt object (50000) Quaoar. *Nature* 432: 731–733.

Johansen A, Youdin AN, and Lithwick Y (2012) Adding particle collisions to the formation of asteroids and Kuiper belt objects via streaming instabilities. *Astronomy & Astrophysics* 537: A125 (17 pp.).

Kaib N, Roškar R, and Quinn T (2011) Sedna and the Oort Cloud around a migrating Sun. *Icarus* 215: 491–507.

Kargel JS (1992) Ammonia-water volcanism on icy satellites: Phase relations at 1 atmosphere. *Icarus* 100: 556–574.

Kenyon SC and Bromley BC (2012) Coagulation calculations of icy planet formation at 15–150 AU: A correlation between the maximum radius and the slope of the size distribution for trans-Neptunian objects. *Astronomical Journal* 143: 63 (21 pp.).

Kenyon SC, Bromley BC, O'Brien DP, and Davis DR (2008) Formation and collisional evolution of Kuiper belt objects. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 293–313. Tucson, AZ: University of Arizona Press.

Lanzerotti LJ, Brown WL, Marcantonio KJ, and Johnson RE (1984) Production of ammonia-depleted surface layers on the saturnian satellites by ion sputtering. *Nature* 312: 139–140.

Leinhardt Z, Marcus RA, and Stewart ST (2010) The formation of the collisional family around the dwarf planet Haumea. *Astrophysical Journal* 714: 1789–1799.

Lellouch E, de Bergh C, Sicardy B, Forget F, Vangvichith M, and Käufl H-U (2014) Exploring the spatial, temporal, and vertical distribution of methane in Pluto's atmosphere. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2014.03.027.

Lellouch E, de Bergh C, Sicardy B, Käufl H-U, Kassi S, and Smette A (2011) High resolution spectroscopy of Pluto's atmosphere: Detection of the 2.3 μm CH_4 bands and evidence for carbon monoxide. *Astronomy & Astrophysics* 530: L1–L4.

Lellouch E, Sicardy B, de Bergh C, Käufl H-U, Kassi S, and Campargue A (2009) Pluto's lower atmosphere structure and methane abundance from high-resolution spectroscopy and stellar occultations. *Astronomy & Astrophysics* 495: L17–L21.

Levison HF, Morbidelli A, VanLaerhoven C, Gomes R, and Tsiganis K (2008) Origin of the structure of the Kuiper belt during a dynamical instability in the orbits of Uranus and Neptune. *Icarus* 196: 258–273.

Levison HF, Morbidelli A, Tsiganis K, Nesvorný D, and Gomes R (2011) Late orbital instabilities in the outer planets induced by interaction with a self-gravitating planetesimal disk. *Astronomical Journal* 142: 152 (11 pp.).

Licandro J, Pinilla-Alonso N, Pedani M, Oliva E, Tozzi GP, and Grundy WM (2006) The methane ice rich surface of large TNO 2005 FY9: A Pluto-twin in the trans-neptunian belt? *Astronomy & Astrophysics* 445: L35–L38.

Lim TL, Stansberry J, Müller TG, et al. (2010) "TNOs are Cool": A survey of the trans-Neptunian region III. Thermophysical properties of 90482 Orcus and 136472 Makemake. *Astronomy & Astrophysics* 518: L148 (5 pp.).

Malhotra R (1995) The origin of Pluto's orbit: Implications for the Solar System beyond Neptune. *Astronomical Journal* 110: 420–429.

McCarthy DW, Hubbard WB, Kulesa CA, et al. (2008) Long-wavelength density fluctuations resolved in Pluto's high atmosphere. *Astronomical Journal* 136: 1519–1522.

McKinnon WB (2002) On the initial thermal evolution of Kuiper belt objects. *Proceedings Asteroids, Comets, Meteors (ACM 2002)* — ESA SP-500, pp. 29–38.

McKinnon WB (2006) On convective instability in the ice I shells of outer solar system bodies, with detailed application to Callisto. *Icarus* 183: 435–450.

McKinnon WB (2013) The shapes of things to come: Rotational and tidal figures of equilibrium for Pluto and Charon. *The Pluto System on the Eve of Exploration by New Horizons: Perspectives and Predictions* (https://dnnpro.outer.jhuapl.edu/plutoscience/Home.aspx).

McKinnon WB, Lunine JI, and Banfield D (1995) Origin and evolution of Triton. In: Cruikshank DP (ed.) *Neptune and Triton*, pp. 807–877. Tucson, AZ: University of Arizona Press.

McKinnon WB, Prialnik D, Stern SA, and Coradini A (2008) Structure and evolution of Kuiper belt objects and dwarf planets. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 213–241. Tucson, AZ: University of Arizona Press.

McKinnon WB, Simonelli D, and Schubert G (1997) Composition, internal structure, and thermal evolution of Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 259–343. Tucson, AZ: University of Arizona Press.

Merlin F, Barucci MA, de Bergh C, et al. (2010) Chemical and physical properties of the variegated Pluto and Charon surfaces. *Icarus* 210: 930–943.

Moore JM, Howard AD, Schenk PM, et al. (2014) Geology before Pluto: Pre-encounter Considerations. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2014.04.028.

Morbidelli A, Bottke WF, Nesvorný D, and Levison HF (2009) Asteroids were born big. *Icarus* 204: 558–573.

Morbidelli A, Gaspar HS, and Nesvorny D (2014) Origin of the peculiar eccentricity distribution of the inner cold Kuiper belt. *Icarus* 232: 81–87.

Morbidelli A, Levison HF, Bottke WF, Dones L, and Nesvorný D (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.

Morbidelli A, Levison HF, and Gomes R (2008) The dynamical structure of the Kuiper belt and its primordial origin. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 275–292. Tucson, AZ: University of Arizona Press.

Mumma MJ and Charnley SB (2011) The chemical composition of comets—Emerging taxonomies and natal heritage. *Annual Review of Astronomy and Astrophysics* 49: 471–524.

Nesvorný D, Vokrouhlický D, and Morbidelli A (2013) Capture of Trojans by jumping Jupiter. *Astrophysical Journal* 768: 45 (8 pp.).

Nesvorný D, Youdin AN, and Richardson DC (2010) Formation of Kuiper belt binaries by gravitational collapse. *Astronomical Journal* 140: 785–793.

Nimmo F and Spencer JR (2014) Powering Triton's recent geological activity by obliquity tides: Implications for Pluto geology. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2014.01.044.

Nogueira E, Brasser R, and Gomes R (2011) Reassessing the origin of Triton. *Icarus* 214: 113–130.

Noll KS, Grundy WM, Chiang EI, Margot J-L, and Kern SD (2008a) Binaries in the Kuiper belt. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 345–363. Tucson, AZ: University of Arizona Press.

Noll KS, Grundy WM, Stephens DC, Levison HF, and Kern SD (2008b) Evidence for two populations of classical transneptunian objects: The strong inclination dependence of classical binaries. *Icarus* 194: 758–768.

Olkin CB, Young LA, Borncamp D, et al. (2014) Evidence that Pluto's atmosphere does not collapse from occultations including the 2013 May 04 event. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2014.3.026.

Olkin CB, Young EF, Young LA, et al. (2007) Pluto's spectrum from 1.0 to 4.2 μm: Implications for surface properties. *Astronomical Journal* 133: 420–431.

Ortiz JL, Thirouin A, Campo Bagatin A, et al. (2012) Rotational fission of trans-Neptunian objects: The case of Haumea. *Monthly Notices of the Royal Astronomical Society* 419: 2315–2324.

Paetzold M, Andert TP, Tyler G, Bird MK, Hinson DP, and Linscott IR (2013) Mass determination of Pluto and Charon from New Horizon REX Radio Science observations. *AAS/DPS Meeting 45*, 310.06.

Pál E, Kiss C, Müller TG, et al. (2012) "TNOs are Cool": A survey of the trans-Neptunian region VII. Size and surface characteristics of (90377) Sedna and 2010 EK_{139}. *Astronomy & Astrophysics* 541: L6 (4 pp.).

Parker AH, Buie MW, Osip DJ, et al. (2013) 2011 HM_{102}: Discovery of a high-inclination L5 Neptune Trojan in the search for a post-Pluto *New Horizons* target. *Astronomical Journal* 145: 96 (6 pp.).

Person MJ, Dunham EW, Bosh AS, et al. (2013) The 2011 June 23 stellar occultation by Pluto: Airborne and ground observations. *Astronomical Journal* 146: 83 (15 pp.).

Person MJ, Elliot JL, Bosh AS, et al. (2010) Pluto's atmosphere from the July 2010 stellar occultation. *AAS/DPS Meeting 42*, 20.03.
Person MJ, Elliot JL, Gulbis AA, et al. (2006) Charon's radius and density from the combined data sets of the 2005 July 11 occultation. *Astronomical Journal* 132: 1575–1580.
Person MJ, Elliot JL, Gulbis AA, et al. (2008) Waves in Pluto's atmosphere. *Astronomical Journal* 136: 1510–1518.
Petit J-M, Kavelaars JJ, Gladman B, and Loredo T (2008) Size distribution of multikilometer transneptunian objects. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 71–87. Tucson, AZ: University of Arizona Press.
Petit J-M, Kavelaars JJ, Gladman BJ, et al. (2011) The Canada-France Ecliptic Plane Survey—Full data release: The orbital structure of the Kuiper belt. *Astronomical Journal* 142: 131–155.
Quirico E, Douté S, Schmitt B, et al. (1999) Composition, physical state, and distribution of ices at the surface of Triton. *Icarus* 139: 159–178.
Robouchon G and Nimmo F (2011) Thermal evolution of Pluto and implications for surface tectonics and a subsurface ocean. *Icarus* 216: 426–439.
Santos-Sanz P, Lellouch E, Fornasier S, et al. (2012) "TNOs are Cool": A survey of the trans-Neptunian region IV. Size/albedo characterization of 15 scattered disk and detached objects observed with *Herschel*-PACS. *Astronomy & Astrophysics* 541: A92 (18 pp.).
Schaller EL and Brown ME (2007a) Detection of methane on Kuiper belt object (50000) Quaoar. *Astrophysical Journal Letters* 670: L49–L51.
Schaller EL and Brown ME (2007b) Volatile loss and retention on Kuiper belt objects. *Astrophysical Journal Letters* 659: L61–L64.
Schenk PM and Zahnle K (2007) On the negligible surface age of Triton. *Icarus* 192: 135–149.
Schlichting HE and Sari R (2009) The creation of Haumea's collisional family. *Astrophysical Journal* 700: 1242–1246.
Schlichting H, Fuentes CI, and Trilling DE (2013) Initial planetesimal sizes and the size distribution of small Kuiper belt objects. *Astronomical Journal* 146: 36 (7 pp.).
Schubert G, Hussmann H, Lainey V, et al. (2010) Evolution of icy satellites. *Space Science Reviews* 153: 447–484. http://dx.doi.org/10.1007/s11214-010-9635-1.
Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.
Schwamb ME, Brown ME, and Fraser WC (2014) The small numbers of large Kuiper belt objects. *Astronomical Journal* 147: 2 (6 pp.).
Shankman C, Gladman BJ, Kaib N, Kavelaars JJ, and Petit JM (2013) A possible divot in the size distribution of the Kuiper belt's scattering objects. *The Astrophysical Journal Letters* 764: L2 (6 pp.).
Sheppard SS (2012) The color differences of Kuiper belt objects in resonance with Neptune. *Astronomical Journal* 144: 169 (14 pp.).
Sheppard SS, Udalski A, Trujillo C, et al. (2011) A southern sky and Galactic Plane survey for bright Kuiper belt objects. *Astronomical Journal* 142: 98 (10 pp.).
Showalter MR, Hamilton DP, Stern SA, Weaver HA, Steffl AJ, and Young LA (2011) New satellite of (134340) Pluto: S/2011 (134340) 1. *CBAT 2769*.
Showalter MR, Weaver HA, Stern SA, et al. (2012) New satellite of (134340) Pluto: S/2012 (134340) 1, *IAU Circular 9253*, 1.
Sicardy B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.
Sicardy B, Bolt G, Broughton J, et al. (2011) Constraints on Charon's orbital elements from the double stellar occultation of 2008 June 22. *Astronomical Journal* 141: 67 (16 pp.).
Sicardy B, Braga-Ribas F, Widemann T, et al. (2012) Charon's size and orbit from double stellar occultations. *AAS/DPS Meeting 44*, 304.01.
Singer KN and McKinnon WB (2011) Tectonics on Iapetus: Despinning, respinning, or something completely different? *Icarus* 216: 198–211.
Sohl F, Choukroun M, and Kargel J (2010) Subsurface water oceans on icy satellites: Chemical composition and exchange processes. *Space Science Reviews* 153: 485–510. http://dx.doi.org/10.1007/s11214-010-9646-y.
Solomatov VS and Moresi L-N (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on Earth and other terrestrial planets. *Journal of Geophysical Research* 105: 21795–21817.
Stansberry J, Grundy W, Brown M, et al. (2008) Physical properties of Kuiper belt and Centaur objects: Constraints from the Spitzer Space Telescope. In: Barucci MA, Boehnhardt H, Cruikshank D, and Morbidelli A (eds.) *The Solar System Beyond Neptune*, pp. 261–279. Tucson, AZ: University of Arizona Press.
Stern SA (2008) The New Horizons Pluto Kuiper Belt Mission: An overview with historical context. *Space Science Reviews* 140: 3–21.
Stern SA and McKinnon WB (2000) Triton's surface image and impactor population revisited in light of Kuiper belt fluxes: Evidence for small Kuiper belt objects and recent geological activity. *Astronomical Journal* 119: 945–952.
Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.
Strazzulla G and Palumbo MA (1998) Evolution of icy surfaces: An experimental approach. *Planetary and Space Science* 46: 1339–1348.
Tegler SC, Cornelison DM, Grundy WM, et al. (2010) Methane and nitrogen abundances on Pluto and Eris. *Astrophysical Journal* 725: 1296–1305.
Tegler SC, Grundy WM, Vilas F, Romanishin W, Cornelison DM, and Consolmagno GJ (2008) Evidence of N_2-ice on the surface of the icy dwarf Planet 136472 (2005 FY9). *Icarus* 195: 844–850.
Tholen DJ and Buie MW (1997) Bulk properties of Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 193–219. Tucson, AZ: University of Arizona Press.
Trujillo C and Sheppard SS (2014) A Sedna-like body with a perihelion of 80 astronomical units. *Nature* 507: 471–474.
Trujillo C, Sheppard SS, and Schaller E (2011) A photometric system for detection of water and methane ices on Kuiper belt objects. *Astrophysical Journal* 730: 105 (13 pp.).
Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Nature* 435: 459–461.
Tyler GL, Linscott IR, Bird MK, et al. (2008) The New Horizons Radio Science Experiment (REX). *Space Science Reviews* 140: 217–259.
Verbiscer AJ, Peterson DE, and Skrutskie MF (2007) Simultaneous spatially-resolved near-infrared spectra of Pluto and Charon. *Lunar and Planetary Science XXXVIII*, #2318.
Vilenius E, Kiss C, Mommert M, et al. (2012) "TNOs are Cool": A survey of the trans-Neptunian region VI. *Herschel*/PACS observations and thermal modeling of 19 classical Kuiper belt objects. *Astronomy & Astrophysics* 541: A94 (17 pp.).
Waite JH, Lewis WS, Magee BA, et al. (2009) Liquid water on Enceladus from observations of ammonia and ^{40}Ar in the plume. *Nature* 460: 487–490. http://dx.doi.org/10.1038/nature08153.
Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) A low mass for Mars from Jupiter's early gas-driven migration. *Nature* 475: 206–209.
Weaver HA, Gibson WC, Tapley MB, Young LA, and Stern SA (2008) Overview of the New Horizons science payload. *Space Science Reviews* 140: 75–91.
Wong MH, Lunine JI, Atreya SK, et al. (2008) Oxygen and other volatiles in the giant planets and their satellites. In: MacPherson GJ, Mittlefehldt DW, and Jones JH (eds.) *Oxygen in Earliest Solar System. Reviews in Mineralogy and Geochemistry*, vol. 68, pp. 219–246.
Yelle RV and Elliot JL (1997) Atmospheric structure and composition. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 347–390. Tucson, AZ: University of Arizona Press.
Yelle RV and Lunine JI (1989) Evidence for a molecule heavier than methane in the atmosphere of Pluto. *Nature* 339: 288–290.
Young LA (2013) Pluto's seasons: New predictions for New Horizons. *The Astrophysics Journal Letters* 766: L22–L28.
Young EF, French RG, Young LA, et al. (2008a) Vertical structure in Pluto's atmosphere from the 2006 June 12 stellar occultation. *Astronomical Journal* 136: 1757–1769.
Young LA, Stern SA, Weaver HA, et al. (2008b) New Horizons: Anticipated scientific investigations at the Pluto system. *Space Science Reviews* 140: 93–127.
Zalucha AM and Gulbis AAS (2012) Comparison of a simple 2-D Pluto general circulation model with stellar occultation light curves and implications for atmospheric circulation. *Journal of Geophysical Research* 117: E05002.
Zalucha AM, Zhu X, Gulbis AAS, Strobel DF, and Elliot JL (2011) An investigation of Pluto's troposphere using stellar occultation light curves and an atmospheric radiative–conductive–convective model. *Icarus* 214: 685–700.
Zhu X, Strobel DF, and Erwin JT (2014) The density and thermal structure of Pluto's atmosphere and associated escape processes and rates. *Icarus* 228: 301–314.

This page intentionally left blank

10.20 Pluto, Charon, and the Kuiper Belt Objects

SA Stern, J Wm Parker, and CB Olkin, Southwest Research Institute, Boulder, CO, USA

© 2007 Elsevier B.V. All rights reserved.

This chapter is reproduced from the previous edition, Volume 10, pp. 541–564, © 2007, Elsevier B.V.

10.20.1	**Overviews**	653
10.20.1.1	The Pluto System	653
10.20.1.2	Kuiper Belt	654
10.20.1.3	Kuiper Belt Objects	655
10.20.1.4	On KBO Nomenclature	655
10.20.2	**Environment**	656
10.20.2.1	Insolation	656
10.20.2.2	Collisions	656
10.20.2.3	Erosional Interactions with the Interstellar Medium	656
10.20.2.4	Radiation Processes	657
10.20.3	**Physical Properties**	657
10.20.3.1	Albedos	657
10.20.3.2	Rotational Properties and Light Curves	658
10.20.3.3	Colors	658
10.20.3.4	Surface Composition	660
10.20.3.5	Atmospheres	662
10.20.3.5.1	Pluto's atmosphere	662
10.20.3.5.2	Other KBOs	664
10.20.3.6	Densities	664
10.20.3.7	Interior Models	665
10.20.4	**Origin**	665
10.20.5	**Future Observational Goals and Prospects**	666
10.20.5.1	Goals	667
10.20.5.2	Future Ground-Based Observatories	667
10.20.5.3	Future Space-Based Observations	667
10.20.5.4	Space Missions to the KB and Centaurs	668
References		668

10.20.1 Overviews

It is now widely recognized that Pluto's discovery was a harbinger of the discovery of the Kuiper Belt (KB), the third major architectural zone of our solar system, after the terrestrial planet and giant planet zones. It is less widely recognized, but nonetheless the case, that Pluto has over time also been a harbinger of dwarf planets, of slow rotation periods, of bodies containing exotic surface volatiles, of bodies trapped in mean motion resonance (MMR) with Neptune, and of satellites that are now seen to be common among Kuiper Belt objects (KBOs). Someday, Pluto may be seen as a harbinger of KBO atmospheres.

In this section, we provide an overview of what is known about the Pluto system, the KB, and KBOs. In subsequent sections, more detail is provided.

10.20.1.1 The Pluto System

The discovery of Pluto by Tombaugh in 1930 opened a chapter in planetary science that is still unfolding. Almost immediately after this discovery was announced, it was recognized that this new planet was highly unusual, owing to its small size and its highly elliptical, highly inclined orbit. As a result, struggles to place Pluto in the context of the architecture of our solar system continued for many decades. The range of conjecture concerning Pluto spanned the space from it being an ejected asteroid, to an escaped satellite, to a member of a distant debris belt related to planet formation (e.g., see Stern et al., 1997).

However, owing to limitations in observing and computing technology, little progress was made in understanding Pluto's context before the 1980s, when serious discussion of the KB (Duncan et al., 1988; Fernandez and Ip, 1984; Stern, 1991) and the population content of 1000 km-sized bodies in the deep outer solar system was seriously examined from a quantitative standpoint.

Similarly, owing to its faintness and small angular size on the sky, which put its study well beyond the capabilities of mid-twentieth-century technology, little progress regarding Pluto's nature was made before the 1950s. However, from 1955 to the present, the increasing pace of technological advances in astronomical telescopes and instrumentation, combined with advances in perspective drawn from the initial reconnaissance of the terrestrial and giant planets, and the discovery of the KB, led to an increasingly full picture of Pluto and its context. Major advances achieved in this era included the following:

- Discovery of Pluto's 6.4-day rotation period (Walker and Hardie, 1955).

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00179-2

- Discovery of Pluto's 3:2 mean motion heliocentric resonance with Neptune (Cohen and Hubbard, 1965).
- First crude spectrum, revealing a reddish surface color (Fix et al., 1970).
- CH_4 ice detected (Cruikshank et al., 1976).
- Charon was discovered, yielding a system mass (Christy and Harrington, 1978).
- Stellar occultation revealed that Charon's radius is close to 600 km (Walker, 1980).
- Onset of a rich series of mutual occultation events between Pluto and Charon (Binzel et al., 1985).
- Determination of reliable radii for Pluto and Charon (Tholen et al., 1986), in turn yielding a 2 g cm^{-3} average system density.
- Water ice was discovered on Charon (Marcialis et al., 1987).
- Pluto's atmosphere was revealed by stellar occultation (Elliot et al., 1989).
- Mutual events produced evidence for polar caps (Binzel, 1989).
- Arguments that ice dwarf planets like Pluto were common in the early solar system (Stern, 1991).
- Discovery of the KB (Jewitt and Luu, 1992).
- Discovery of N_2 and CO ices on Pluto (Owen et al., 1993).
- Hubble Space Telescope (HST) imaging revealed Pluto's highly variegated surface (Stern et al., 1997).
- Detection of CH_4 in Pluto's atmosphere (Young et al., 1997).
- Discovery of Pluto's small satellites Nix and Hydra (Stern et al., 2006; Weaver et al., 2006).

With many of these facts in hand, by the late 1980s, it was clear that Pluto was a very different kind of world from either the terrestrial planets or the giant planets, replete with multiple surface volatiles; a complex, escaping, seasonally variable atmosphere; and a highly variegated surface. Moreover, Pluto was seen to be a binary planet, that is, an object with a satellite so large, and so close, that the system barycenter lies outside the primary in free space between the planet and the moon. By the early 1990s, it had also been argued that Pluto was not a misfit among the planets, but the first of a large population of 1000 km-sized KB and more distant bodies in a population that likely include objects of sizes extending up to Earth mass or somewhat larger (Stern, 1991).

10.20.1.2 Kuiper Belt

The existence of the KB was first predicted by mid-twentieth-century astronomers such as Kenneth Edgeworth (1943, 1949) and Gerard Kuiper (1951). These and other astronomers of the 1930s, 1940s, and 1950s postulated that a debris belt of material left over from planetary formation might orbit the Sun beyond Neptune. However, the faint object detection technology of the mid-twentieth century was too primitive to give astronomers real hope of finding bodies out there – they were simply too faint to be found.

By the late 1980s, however, dynamicists found strong evidence in the inclination distribution of the Jupiter family comets that they are coming from a disklike reservoir just beyond Neptune's orbit (Duncan et al., 1988). This dynamic evidence was bolstered by the discovery of 2060 Chiron, orbiting in a short-lived, dynamically chaotic orbit, between Saturn and Uranus (Kowal et al., 1977). Unfortunately, although Kowal's survey data set did detect several large KBOs, his manual detection technique failed to identify them as trans-Neptunian bodies, so they remained undiscovered for another 20+ years.

As a result of the discovery of Chiron and computer modeling that argued strongly for a disklike reservoir of comets beyond the giant planets (e.g., Duncan et al., 1987), a number of searches were begun in the late 1980s for the belt of material that Kuiper predicted. The first KBO was subsequently discovered (Jewitt and Luu, 1992). This object, designated 1992 QB_1, is more than 1000 times fainter than Pluto and probably some 10–20 times smaller in radius.

Over 1200 other KBOs and centaurs had been discovered by late 2006, with estimated diameters ranging from 30 to perhaps 2700 km (e.g., Bernstein et al., 2004; Bertoldi et al., 2006). Based on the large fraction of sky left to be searched and the sky density of faint, distant objects being found in deep charge-coupled device (CCD) images, it is estimated that over 100 000 KBOs with diameters more than 50 km may orbit in the KB. From the orbital distribution of KBOs, the KB has been revealed to be a disklike or belt-like structure that stretches from 30 to at least 55 astronomical units (AU) from the Sun. It is now recognized that the KB is a far greater collection of objects, particularly large objects, than the asteroid belt between Mars and Jupiter.

Collisional processes are known to play a key role in the KB (e.g., Davis and Farinella, 1997; Durda and Stern, 2000; Stern, 1995, 1996). A significant result of collisional modeling is that KBOs smaller than ~50 km in diameter cannot have survived the collisional bombardment of the current-day KB over time and therefore must be much younger than the age of the solar system. This implies that the Jupiter family comets, which have their source region in the KB, are recently produced from chipped-off larger, ancient KBOs.

Based on the sizes and orbits of KBOs, it appears that the birthplace of the KBOs was well on its way to growing one or more large planets, perhaps even something the size of the Earth, or even Neptune, when the growth process was interrupted (e.g., Farinella et al., 2000; Stern and Colwell, 1997a,b). Perhaps, the formation or transport of Neptune near 30 AU interrupted this growth.

Numerous excellent reviews of the structure and dynamics of the KB have been written (e.g., Morbidelli, 2006). Here, we give only a brief recapitulation of the primary structure and orbital populations of the KB, in order to set the context for our discussions concerning KBOs themselves.

The KB consists of two broad dynamic populations, the classical KB (CKB) and the scattered KB (SKB). Both populations involve bodies on orbits that lie beyond Neptune. Former KBOs, which currently penetrate deeply into the giant planets' region owing to dynamic loss from the CKB and SKB, are referred to as centaurs. Centaur bodies have short (e.g., $10^{6.5}$–$10^{7.5}$ years) orbital lifetimes that typically end with a strong scattering event that ejects them back outside the giant planet region.

The classical belt contains objects that orbit beyond Neptune but generally within ~60 AU. Both Neptune-resonant and Neptune-nonresonant classical KBOs (CKBOs) are recognized. Those on resonant orbits are often referred to as

plutinos, in honor of Pluto, the first object discovered on such an orbit type. Although Pluto orbits in the 3:2 Neptune MMR, plutinos have also been found in the 2:1, 5:3, and 7:4 Neptune MMRs as well.

The MMR orbits of the plutinos provide strong evidence for the migration of Neptune, probably by many astronomical units, early in the history of the solar system (e.g., Malhotra, 1993; Morbidelli et al., 2005). Further, the broad distribution of CKBO orbital inclinations and the presence of the scattered KBOs (SKBOs) indicate a violent dynamic history there. This might involve a passing star, large migrating planetary embryos, or sweeping secular resonances during the time that the KBOs were forming (see, e.g., Morbidelli (2006)).

Estimates of the mass of the current-day KB fall in the range of $10^{-2}\,M_{\rm Earth}$ to $10^{-1}\,M_{\rm Earth}$, with crudely equal proportions in the CKB and SKB (Morbidelli, 2006). Less than 0.1% of this mass is at any time orbiting closer as centaurs (Morbidelli, 2006).

The fact that large bodies hundreds and thousands of kilometers across orbit in the KB provides inescapable evidence that these bodies formed in an environment with approximately 10^2–10^3 times higher surface mass density than is present today (e.g., Farinella et al., 2000; Kenyon and Bromley, 2001; Stern, 1995, 1996). This in turn suggests either that the KB region was formerly far more massive and was subsequently dynamically and/or collisionally eroded (e.g., Stern and Colwell, 1997a,b) or that the large bodies in the KB were dynamically transported outward to their current orbits after formation closer to the Sun (Morbidelli, 2006).

A key dynamic characteristic of the KB is that many of the orbits there are dynamically 'hot,' that is, they have substantial eccentricities and inclinations as high as 40°. These orbits provide strong forensic evidence of one or more dynamic excitation mechanisms that operated in the past. Suggested excitation mechanisms include but are not limited to scattering by the giant planets, migration of the giant planets, perturbations from formerly present planets in the KB, and external perturbations from passing stars. Hot orbits are also common to the SKB. Most discovered SKBOs were very likely excited to their current orbits by scattering off Neptune.

The CKB also contains a population of bodies on cold (low-eccentricity and low-inclination) nonresonant orbits (Brown, 2001; Levison and Stern, 2001) that imply formation in situ, that is, without the strong dynamic transport of the hot population. This population includes large KBOs, providing strong evidence that there was indeed sufficient mass in the primordial disk extension beyond Neptune for such objects to form.

A cold SKB population has also been proposed (Stern, 1996; Stern and Colwell, 1997a) in order to explain the surface mass density deficit of the KB and the possible in situ growth of Sedna (Stern, 2005). Because no such bodies have yet been detected, it is not yet possible to determine definitively whether the primordial protoplanetary disk was compact (i.e., ~20–30 AU in radius), a less compact structure (perhaps 50 AU in radius), or a broad disk (which might reach out to 500 AU or more, as observed in many extrasolar debris disks).

10.20.1.3 Kuiper Belt Objects

The properties of KBOs are discussed in greater detail in the succeeding text, but a few relevant facts are as follows. The surfaces of KBOs are relatively unreflective, typically scattering only 3% to perhaps 15% of the light that falls on them, though Eris and Pluto have very high surface albedos (e.g., Grundy et al., 2005a,b). KBOs have a wide range of surface colors (Barucci et al., 1999a,b; Davies, 2001; Jewitt and Luu, 1998; Luu and Jewitt, 1996; Tegler and Romanishin, 1998), varying from almost gray to very red, but it is not clear whether this is due to genetic differences among KBOs or evolutionary effects (e.g., space weathering and collisional resurfacing). It is thought that KBOs consist primarily of mixtures of water ice and rock, with some amount of organic and other complex compounds as well. Most KBOs rotate on their axes in a few hours, but some take days to rotate. Some have sufficiently low enough densities ($<1\ {\rm g\ cm^{-3}}$) to clearly indicate internal macroporosity (e.g., McKinnon, 2002; Mckinnon and Barr, 2006). There is already evidence for water and more exotic ices like N_2 and CH_4 on KBOs (e.g., Delsanti and Jewitt, 2006; Licandro et al., 2006). But it is not known if KBOs fall into clear compositional groups as the asteroids do or whether they instead simply form a broad continuum of surface colors and compositions. Other than Pluto, no KBO is known to have an atmosphere, but at last, there are good reasons to expect that some large KBOs exhibit atmospheres, at least transiently (Stern and Trafton, in press).

In mid-2005, it was also discovered that Pluto has two smaller moons (Weaver et al., 2006), making the system a quadruple. Pluto's smaller satellites orbit in nearly circular orbits and in the same plane as Charon; they lie near or at the location of MMRs with Charon. These attributes strongly imply a common origin for all of Pluto's satellites in the giant collision that created Charon (Stern et al., 2006).

In 2001, the first satellite orbiting a KBO, 1998 WW31, was discovered (Veillet et al., 2002). Since that time, numerous other KBOs have been discovered to have satellites. Based on detection statistics and taking into account selection effects that bias detection frequency, it has recently been estimated that at least 10% of all KBOs have satellites (Margot et al., 2004; Noll et al., 2002; Stephens and Noll, 2005); owing to some strong detection biases, the actual fraction of KBOs with satellites may be considerably higher. By late 2006, within 5 years after the discovery of the first KBO binary, there were over two dozen KBOs and centaurs with known satellites. No KBOs except Pluto (with three known satellites) and 2003 EL61 (with two known satellites) are as yet known to have more than one satellite.

Knowledge about KBO satellite albedos is nonexistent; so are colors, except for Pluto's three satellites (Stern et al., 2006). As to satellite spectra, the only published information we are aware of as of mid-2006 is that the brightest satellite of 2003 EL61 exhibits strong water ice features at 1.5 and 2.0 μm that are much deeper than typically found on KBOs (Barkume et al., 2006). We hope that much more information will soon be obtained on KBO satellites, since interesting questions revolve around whether KBOs and their satellites are similar in color, composition, and albedo, or not. Whether the highly dichotomous colors, surface compositions, and albedos of Pluto and Charon are typical, or not, is an open question.

10.20.1.4 On KBO Nomenclature

We, like many others, refer to all bodies orbiting in the KB as KBOs. Therefore, in effect, the KBO designator is merely an

indication of the location of bodies orbiting there, rather than an indication of the kinds of bodies orbiting there. The term 'trans-Neptunian object' (TNO) also is used in the literature, and we occasionally use it in what follows, also as an indication of location without any information content about the kind of body that object in this region is.

It is also useful to classify KBOs into size categories as follows:

- *Debris*. KBOs with diameters $D < 50$ m
- *Comets*. KBOs with 50 m $< D <$ 500 km
- *Dwarf planets*. KBOs with 500 km $< D <$ 5000 km
- *Satellites*. Secondary bodies orbiting any KBO

If KB bodies with diameters >5000 km are later discovered (the SKB region is the most likely region in this regard), we would refer to these objects as planets.

In what follows, we will also use the term 'binary' from time to time. When we do so, we are referring to two objects that orbit a common center of mass and that have a barycenter in free space between the two orbiting bodies. An example of a KBO binary is thus the binary planet Pluto–Charon.

10.20.2 Environment

The salient environmental characteristics of the KB region are its low population and surface mass density, its cold thermal conditions, and its location within the solar heliosphere. We now discuss the various environmental factors that influence KBOs; these include insolation, collisions, ISM grain erosion effects, and radiation fluence. We describe each in turn. These various influences affect the surfaces of KBOs and therefore shed light on the interpretation of KBO physical measurements, which are discussed later in this chapter.

10.20.2.1 Insolation

The thermal environment of the KB is dominated by the Sun. Although the extant insolation is diluted relative to conditions at 1 AU by factors of order 10^3–10^4, it is still capable of heating surfaces to characteristic temperatures of ~20 to ~70 K, depending on heliocentric distance, polar obliquity, surface albedo, and surface emissivity.

Because many KB bodies range over great swaths of heliocentric distance (as Pluto does) and have substantial polar obliquities (as Pluto also does), significant annual and seasonal variations in surface temperature are expected on individual KBOs.

At the cold temperatures characteristic of the KB, virtually all cosmogonically abundant molecules have negligible vapor pressures; this includes the common outer solar system ices H_2O and CO_2. The only thermophysically active molecules of cosmogonic importance in the KB temperature regime are, in order of decreasing volatility with temperature, O_2, N_2, CO, and CH_4. In addition to these four molecular species, the noble gases He, Ne, Ar, and Kr are also volatile in the KB thermal regime. For those KBOs with surface-lying or near-surface-lying volatiles, one expects insolation to drive sublimation, followed by the formation of either global or patchy atmospheres (Stern and Trafton, in press).

10.20.2.2 Collisions

As discussed previously, collisional processes are important in the KB; this has been known for some time (e.g., Stern, 1992, 1995; see Farinella et al. (2000) for a thorough review). Although, owing to the more dilute population density and lower orbital speeds, intrinsic collision rates (the number of collisions per unit surface area per unit time) are lower by a factor of about 1000 in the KB compared to the main asteroid belt, the population of objects there is roughly 1000 times as great. As a result, the overall level of collisional processing of individual objects is of similar scale in the KB as in the main asteroid belt.

Durda and Stern (2000) showed that in the current KB, surface modification by small impactors is expected to be extensive. In particular, they estimated that the cumulative fraction of the surface area of KBOs ranges from a few to a few tens of percent over 3.5 Gy. Such collisions structurally modify surfaces and excavate buried materials, potentially including volatiles, from depths up to several times the impactor diameter.

At typical KB collision speeds, typically 0.5–1.5 km s^{-1}, little surface vaporization of H_2O and CO_2 is expected. What about vaporization of more volatile surface ices? Unfortunately, no studies of the vaporization fraction of more volatiles ices like N_2, CO, and CH_4 are available, so it is not known whether significant mobilization or even loss of these volatiles occurs (E. Pierrazzo, personal communication). However, by excavating fresh material, collisions may be responsible for creating both high-albedo ejecta rays that influence KBO albedos, and the competition between the excavation of fresh material and its subsequent reddening by radiation may also create surface color variability with time (e.g., Jewitt, 2002; Stern, 2003).

KB collisions happen at sufficiently high speeds that they are erosional for bodies with diameters of a few hundred kilometers and smaller, causing KBO comets and smaller debris to lose surface material over time. In fact, the time-averaged rate of erosion in the KB is so high that virtually all KBOs with diameters of 5 km and smaller are thought to be fragments from 'chipped-off' larger KBOs in the last 10–20% of the age of the solar system (e.g., Davis and Farinella, 1997; Durda and Stern, 2000; Stern, 1996). Hence, most KB comets are expected to exhibit low surface exposure ages and are therefore not expected to exhibit many impact craters.

10.20.2.3 Erosional Interactions with the Interstellar Medium

The Sun carries the KB through the interstellar medium (ISM) at a high velocity of approximately 30 km s^{-1}. The resulting impacts of KBOs with ISM grain are important in several ways. They are a contributor of darkening agents to KBO surfaces. They can also erode KBO surfaces, particularly on bodies too small (i.e., with <500 km diameter) for the resulting ejecta to remain bound to the target. This process may in total have caused KBOs with sizes of 10 to perhaps 200 km to have lost from about 1 m to perhaps 20 m of surface material since their formation (Stern, 2003).

10.20.2.4 Radiation Processes

The radiation environment of the KB naturally divides into two subcategories: energetic photon bombardment and charged particle bombardment. Interstellar and solar ultraviolet (UV; i.e., $h\nu > 3$ eV) photons have copious fluxes in the KB, providing the energy necessary to break bonds and initiate substantial chemical change in Oort cloud comet surfaces. Because of their much closer proximity to the Sun, KBOs experience a much (i.e., $\sim 10^5$ times) higher UV and solar cosmic ray (SCR) surface doses, dramatically increasing the total deposited charged particle energy incident on the surfaces of these bodies, relative to Oort cloud comets. However, the galactic cosmic ray (GCR) dose on KBOs inside roughly 100 AU is expected to be far reduced, relative to Oort cloud comets, because of the shielding effects of the heliosphere.

UV photosputtering is capable of eroding away the uppermost few micrometers of icy surfaces of atmosphereless bodies (Throop, 2000). But more importantly, in a classic series of laboratory experiments and theoretical studies, UV photons would produce significant alteration of the composition, color, and volatility of the upper several to few tens of micrometers of cometary surfaces (e.g., Greenberg, 1982). Thompson et al. (1987) and Hudson and Moore (2001) confirmed and extended these results, showing that UV photons promote surface darkening (to albedos of only a few percent) and a devolatilization that becomes progressively more severe with dosage and therefore age.

Like UV photons, charged particle radiation is capable of both sputtering surfaces and breaking bonds, thereby inducing chemical reactions and consequently reordering the surface ice matrix (e.g., Gerakines et al., 2001). The irreversible radiation-driven conversion of the original, water-dominated ice matrix to a more complex 'crust' inevitably leads to the darkening of surfaces, via the formation of longer-chain molecular species, notably hydrocarbons and nitriles. The cosmic ray damage layer on KBOs may reach several meters in depth, depending on the near-surface density and porosity (e.g., Strazzulla and Johnson, 1991).

The red color of Pluto's surface is widely believed to be the result of radiolysis on its CH_4—N_2—CO surface. The red colors of many KBOs (see succeeding text) may also be the result of radiolysis, though there may also be intrinsically red materials on these bodies as well.

10.20.3 Physical Properties

In this section, we summarize the current state of knowledge about the physical properties of KBOs. We begin with their albedos, light curves, and colors. We then discuss KBO surface compositions and atmospheres. We then discuss KBO densities and internal structure. In each area discussed, we compare what is known about the Pluto system to other KBOs.

10.20.3.1 Albedos

Owing to their small sizes on the sky and photometric faintness, few KBOs have measured albedos. The 'albedo' of an object is a measure of its reflectivity (the ratio of reflected to incident light). Since the albedo can be a function of wavelength, it is common to refer to the wavelength band or magnitude region of the measurement (e.g., the 'V-magnitude albedo' refers to the reflectivity of the object in the wavelength region covered by the commonly used astronomical V (green–red) filter).

Not surprisingly, Pluto and Charon's albedos are best known. As discussed by Tholen and Buie (1997), Pluto itself has a V-magnitude albedo of 0.55; Charon's V-albedo is near 0.37.

Initially, many workers assumed that KBOs would have comet-like albedos of 0.04. And indeed, when the first measurements of KBO sizes and albedos were obtained, for two relatively large KBOs (1993 SC and 1996 TL66), using infrared (IR) measurements from the ESA Infrared Space Observatory (ISO), results gave radii of 164 km, corresponding to albedos of 0.02–0.03 (Thomas et al., 2000). Soon, however, other KBOs and centaurs were observed and were found with much higher albedos; the centaurs Chiron and Asbolus were found to exhibit albedos of 0.17 ± 0.02 and 0.12 ± 0.03, respectively (Fernández et al., 2002).

Circumstantial evidence for higher albedos also came from KBO satellite formation studies (Stern, 2003), which indicated the need for typical albedos of 0.15 or higher in order to make collisions sufficiently common to explain the growing population of KBO binaries. Then, KBO 2002 AW197 was found to have an albedo of 0.17 ± 0.03 (Cruikshank et al., 2005), and Lykawka and Mukai (2005) and Grundy et al. (2005a,b) demonstrated a tendency for increasingly larger objects to have increasingly higher albedos (**Figure 1**).

Table 1 reveals that few measured KBO albedos are <5%, and few are higher than 20%; instead, most measured albedos lie in the range 15 ± 7%. The few KBOs with very high albedos, like Pluto and Eris, may – like Pluto – exhibit atmospheric recycling mechanisms that refresh the surface albedo (e.g., Stern et al., 1988).

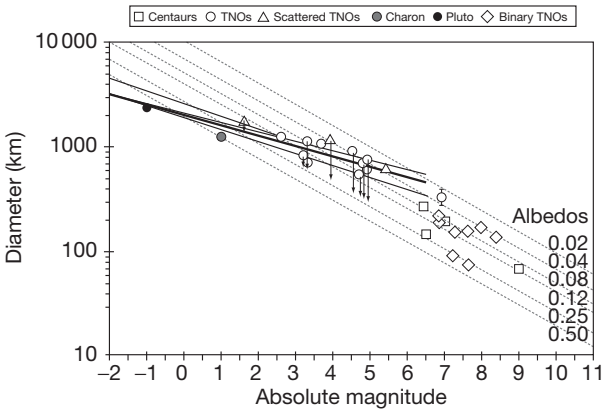


Figure 1 Some KBO (aka TNO) and centaur albedos determined from measured sizes and absolute magnitudes. The dashed line shows various albedo values that can be assumed when size measurements are not available. Reproduced from Lykawka PS and Mukai T (2005) Higher albedos and size distribution of large trans-Neptunian objects. *Planetary and Space Science* 53: 1319–1330.

Table 1 Measured V- and R-magnitude albedos of KBOs

Object	*Dynamic class*	*Albedo (%)*	*Method*	*References*
(58534) 1997 CQ29	Classical	39 ± 17	Binary orbit	G2005
(19521) Chaos	Classical	≥ 5.8	Thermal (1.2 mm)	G2005
1998 WW31	Classical	6.0 ± 2.6	Binary orbit	G2005
(66652) 1999 RZ253	Classical	29 ± 12	Binary orbit	G2005
(20000) Varuna	Classical	$3.7^{+1.1}_{-1.4}$	Thermal (0.9, 1.2 mm)	G2005
(88611)2001 QT297	Classical	10 ± 4	Binary orbit	G2005
(50000) Quaoar	Classical	10 ± 3	Direct imaging	G2005
2003 EL61	Classical	≥ 60	Binary orbit	R2006
(26308) 1998 SM165	2:1	14 ± 6	Binary orbit	G2005
(15789) 1993 SC	3:2	$3.5^{+1.6}_{-1.3}$	Thermal (90 μm)	G2005
(19308) 1996 TO66	3:2	≥ 3.3	Thermal (1.2 mm)	G2005
(47171) 1999 TC36	3:2	22 ± 10	Binary orbit	G2005
		$7.9^{+3.1}_{-2.4}$	Thermal (24,70 μm)	S2006
(28978) Ixion	3:2	≥ 14.8	Thermal (1.2 mm)	G2005
(24835) 1995 SM55	Scattered	≥ 6.7	Thermal (1.2 mm)	G2005
(15874) 1996 TL66	Scattered	≥ 1.8	Thermal (90 μm)	G2005
(38628) Huya	Scattered	≥ 8.4	Thermal (1.2 mm)	G2005
2001 QC298	Scattered	2.5 ± 1.1	Binary orbit	G2005
(55565) 2002 AW197	Scattered	$10.1^{+3.8}_{-2.2}$	Thermal (1.2 mm)	G2005
		17 ± 3	Thermal (24, 70 mm)	C2005
(55636) 2002 TX300	Scattered	≥ 19	Thermal (1.2 mm)	G2005
(84522) 2002 TC302	Scattered	5.1	Thermal (1.2 mm)	G2005
(90377) Sedna	Scattered	4.6	Thermal and imaging	G2005

C2005: Cruikshank et al. (2005) V-band; G2005: Grundy et al. (2005a,b) R-band; R2006: Rabinowitz et al. (2006); S2006: Stansberry et al. (2006) V-band.

It will be interesting to see how increasingly larger albedo data sets affect these conclusions and whether distinctions appear with dynamic class.

10.20.3.2 Rotational Properties and Light Curves

Rotation rates and light curves of KBOs hold clues to their shapes and surface variegation, as well as to their rotational dynamics. The definitive detection of Pluto's rotation rate came from photoelectrically observing its 6.4-day rotational light curve over 50 years ago (Walker and Hardie, 1955). Recently, it has been possible to use HST to separate the individual light from Pluto and Charon (Buie et al., 1997; see **Figure 2**), showing rotation rates and orbital period of the pair are synchronous, which is consistent with expectations from tidal evolution theory (e.g., Dobrovolskis et al., 1997). Of more direct importance to our interests here, Charon's low-amplitude (~8% peak-to-peak) light curve was shown to exhibit little variation, quite unlike Pluto's high-amplitude (~30% peak-to-peak) light curve.

Over 65 KBOs and centaurs now have measured light curves or light curve upper limits (Trilling and Bernstein, 2006), as shown in **Figure 3**. Approximately half of these have established light curve periods and amplitudes. All measurements of KB light curve periods except Pluto's lie between 3.5 and 27 h. Only three exceed 16 h. Hence, to first order, KBO light curve periods appear to be similar to asteroid light curve periods. Most measured KBO light curve amplitudes are <25%. It is not known what relative roles shape effects and albedo effects play in the observed population of measured light curve amplitudes.

10.20.3.3 Colors

The faintness of KBOs makes it difficult to measure accurate photometry and colors. This is problematic in several respects, including the biasing of available measurements to brighter, and therefore typically larger, KBOs.

As with rotation albedos and periods, the first color information obtained on objects in the KB was from measurements of Pluto. Pluto's surface was determined to be red as early as the 1950s. Later, color variations were observed with rotational phase, and color maps were made from HST (Stern et al., 1997) and mutual event data (e.g., Binzel, 1988; Young et al., 2001). In contrast to Pluto, Charon's surface appears to be spectrally neutral, with no significant rotational variation.

The first collections of KBO color data were published in the late 1990s (see Barucci et al., 1999a,b; Davies, 2001; Jewitt and Luu, 1998; Luu and Jewitt, 1996; Tegler and Romanishin, 1998). However, whereas Luu and Jewitt (1996) found a continuous range among the colors of their 13 objects (four centaurs and nine KBOs), Tegler and Romanishin (1998) found that KBO colors fell into two distinct populations separated by a color gap: one population had colors slightly redder than the Sun, and the other had colors significantly redder, containing the reddest objects in the solar system.

This disagreement spurred observations by several groups, none of whom found evidence for a bimodal color distribution (e.g., Barucci et al., 2000a,b; Boehnhardt et al., 2001; Delsanti et al., 2001). A statistical, principal component analysis of 22 KBOs (Barucci et al., 2001) also found that the colors of KBOs and centaurs were consistent with a continuous (rather than a bimodal) distribution, and when Jewitt and Luu (2001) analyzed a larger data set than in their 1996 and 1998 papers,

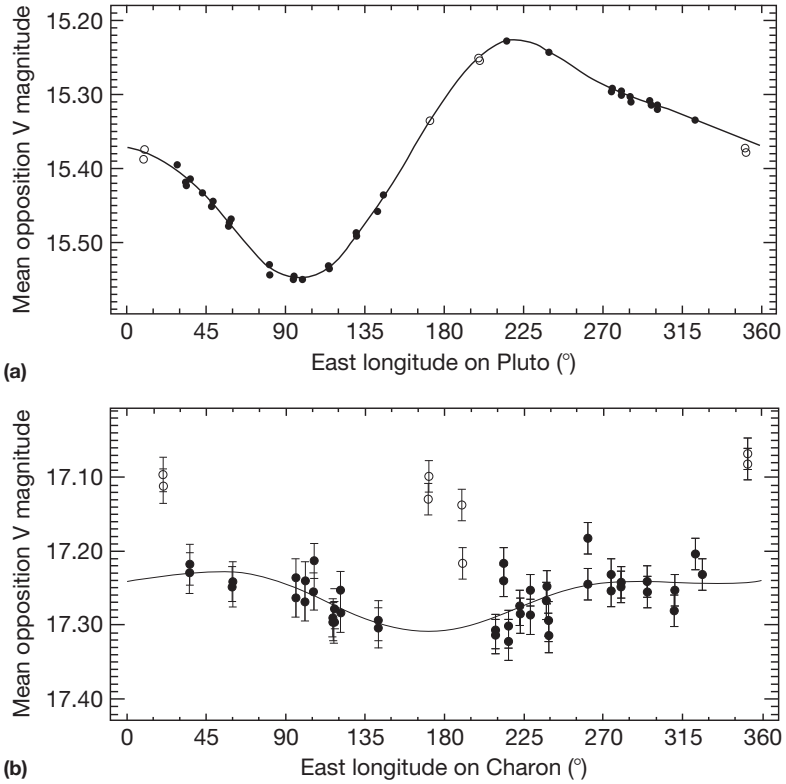


Figure 2 Rotational light curves of Pluto (a) and Charon (b). The mean opposition V magnitude is plotted against sub-Earth longitude in degrees. For Pluto, the error bars are smaller than the symbols and are therefore not shown. The open circles were not included in the fit and correspond to points in the orbit when Pluto and Charon are at a minimum separation. The solid curves represent best-fit light curves. Reproduced from Buie MW, Tholen DJ, and Wasserman LH (1997) Separate light curves of Pluto and Charon. *Icarus* 125: 233–244.

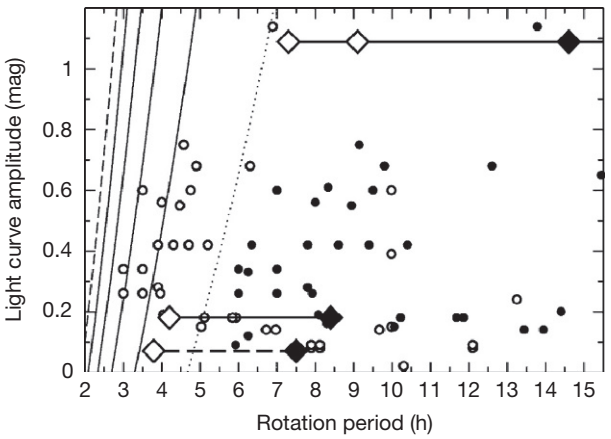


Figure 3 Measured light curve amplitudes of KBOs and centaurs (Trilling and Bernstein, 2006 – TB06). Diamonds represent TB06's own data; circles represent data from other sources. Open symbols indicate photometric periods, and filled symbols indicate rotation periods that are twice the observed photometric period. Objects known to have double-peaked light curves are plotted only as filled symbols. The curves represent solutions for critical rotation periods for various densities in grams per cubic centimeter: 0.5 (dotted line), 1, 1.5, 2, 2.5, and 3 (dashed line). Regions above and to the left of a given curve are dynamically unstable for a given density.

they again found a continuous distribution. However, when Tegler and Romanishin (2003) performed a statistical analysis of 107 KBOs (50 objects from their own observations and 57 objects observed by other groups), they again found a bimodal distribution, which they argued is not detectable in the measurements of other groups due to larger average photometric uncertainties. Peixinho et al. (2003) countered that the bimodal distribution exists within the centaurs, but not KBOs themselves (**Figure 4**).

Currently, there is a broad consensus that KBOs and centaurs display a wide range of colors; but it is not clear whether the color distribution is continuous or separated into distinct groups, though the majority of papers claim the former.

What is the reason for the wide range of KBO colors? Is it related to their origin (e.g., composition) or their evolution (e.g., owing to impacts, transient atmospheres, and varying radiation environments)? Alterations by impacts can compete with irradiation by cosmic rays, so the color of objects in different regions could vary with the local rate of impacts (Gil-Hutton, 2002).

To attempt to constrain the origin versus evolution scenarios further, a variety of workers have searched for color correlations with various orbital and physical parameters in order to address the source of the KBO color diversity. Tegler and Romanishin (1998) did not find any correlation of their

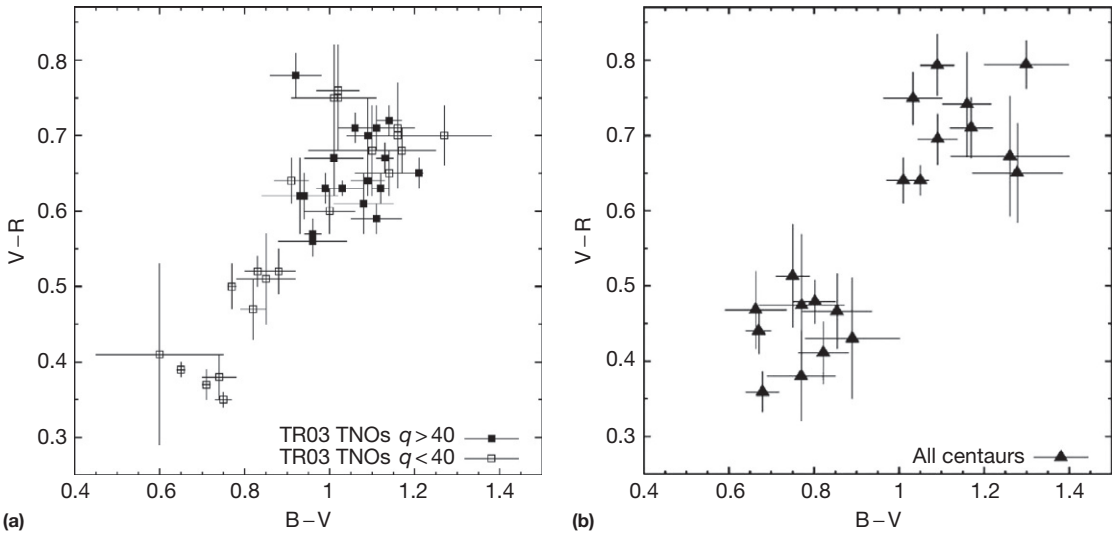


Figure 4 V–R and B–V magnitude color–color plot for KBOs (a) and centaurs (b). B–V is the brightness difference (hence color) seen through the commonly used 'blue' and 'visible' standard astronomical filters, and V–R is the color between the 'visible' and 'red' filters. Reproduced from Peixinho N, Doressoundiram A, Delsanti A, Boehnhardt H, Barucci MA, and Belskaya I (2003) Reopening the TNOs color controversy: Centaurs bimodality and TNOs unimodality. *Astronomy and Astrophysics* 410: L29–L32.

color populations with orbital parameters of the objects, speculating that other factors in the objects' history must come into play, such as solar UV photons, solar-wind and solar-flare particles, and cosmic rays. Barucci et al. (1999a,b) not only discussed the compositional surface variety among centaurs but also found no relationship with orbital parameters, for example, perihelion distance. Tegler and Romanishin (2000) however reported that all nine KBOs in their survey with nearly circular orbits with perihelion distances greater than 40 AU have very red surfaces. This work was later extended to show that the color variations with different dynamic classes point toward a primordial origin, with gray objects forming closer to the Sun than red objects. Hainaut and Delsanti (2002) found that cold classical belt objects apparently have significantly redder colors, whereas the plutinos have the opposite trend. However, Trujillo and Brown (2002) found a correlation between the BVR colors and the orbital inclination of classical KBOs. McBride et al. (2003) found a weak correlation between inclination and color in the classical KB, which they more specifically characterize as a lack of blue objects with low inclinations. Stern (2002) found a correlation between colors of KBOs and the mean random impact speed the objects would experience as a function of their orbital eccentricities and inclinations. This is consistent with KBO colors being driven by a competition between collisional excavation of neutral material and subsequent radiation reddening. Thébault and Doressoundiram (2003) and Doressoundiram et al. (2005) also found a correlation of color with mean impact velocity.

At present, the root cause or causes of the color variation among KBOs remain ambiguous. There is clearly no shortage of ideas; instead, however, there is a shortage of data from which to make strong conclusions. It may require several hundred or several thousand KBOs with well-established colors for definitive progress to be made. Progress will also derive from studies of color variegation on KBOs as they rotate and with the eventual, hoped-for return of color imagery of KBOs by *New Horizons*.

10.20.3.4 Surface Composition

As with other types of physical studies, the spectra of objects in the KB were taken on Pluto. Methane frost was identified on Pluto's surface by Cruikshank et al. (1976). Owen et al. (1993) later added the identification of N_2 and CO to the inventory of Pluto surface frosts. Owen et al. (1993) also recognized that methane must be mixed at the molecular level with nitrogen because of the shift in the methane band centers. Later, Douté et al. (1999) identified more bands of CO and CH_4 on Pluto's surface, in an analysis of IR spectra from 1.4 to 2.5 μm (**Figure 5**). Recently, Cruikshank et al. (2006) reported the detection of C_2H_6 (ethane) on Pluto's surface.

Studies by Douté et al. (1999), Grundy and Fink (1996), and Grundy and Buie (2001) provided important information of the compositional heterogeneity of Pluto's surface (**Figure 6**). They found that CO frost is more abundant in the peak in the sub-Charon hemisphere and that two distinct units containing CH_4 frost exist: one described by the strong methane bands that correlates with the visible light curve and another described by the weak methane bands that does not correlate with the visible light curve.

Pluto's surface variegation in albedo, color, and composition is likely due to a combination of internal activity, volatile transport, and atmosphere–surface interactions; impacts may also play a role. At present, the available data sets do not have sufficient resolution to distinguish uniquely between the various possibilities in order to determine which aspects of the variegation at which locations are due to specific processes.

Owing to Charon's close angular proximity to Pluto, spectroscopy of Charon itself was first possible only as a result of mutual event observations. Such spectra were obtained in the visible by Fink and DiSanti (1988) and in the near-IR by Marcialis et al. (1987) and Buie et al. (1987). This technique involved subtracting the Pluto-only spectrum (when Charon was behind Pluto) from the combined spectrum taken just

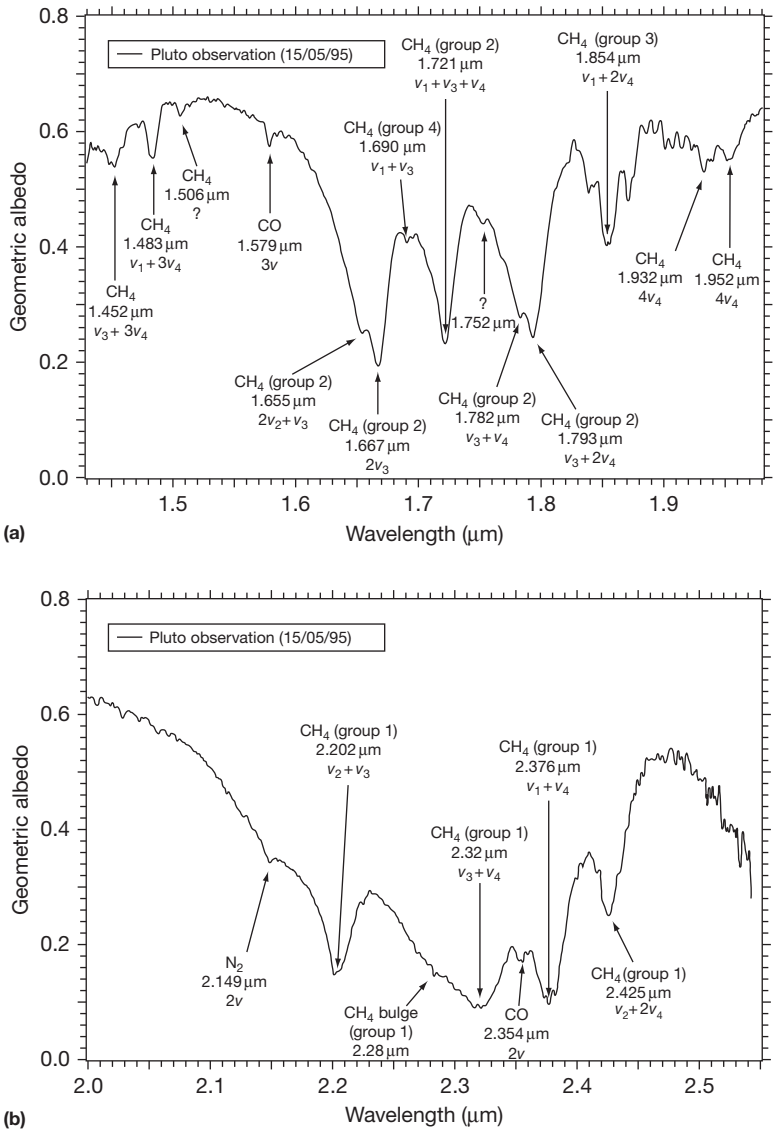

Pluto observation (15/05/95)
Geometric albedo
Wavelength (μm)
(a)
CH4 1.452 μm v3 + 3v4
CH4 1.483 μm v1 + 3v4
CH4 1.506 μm ?
CO 1.579 μm 3v
CH4 (group 2) 1.655 μm 2v2 + v3
CH4 (group 2) 1.667 μm 2v3
CH4 (group 4) 1.690 μm v1 + v3
CH4 (group 2) 1.721 μm v1 + v3 + v4
? 1.752 μm
CH4 (group 2) 1.782 μm v3 + v4
CH4 (group 2) 1.793 μm v3 + 2v4
CH4 (group 3) 1.854 μm v1 + 2v4
CH4 1.932 μm 4v4
CH4 1.952 μm 4v4
Pluto observation (15/05/95)
Geometric albedo
Wavelength (μm)
(b)
N2 2.149 μm 2v
CH4 (group 1) 2.202 μm v2 + v3
CH4 bulge (group 1) 2.28 μm
CH4 (group 1) 2.32 μm v3 + v4
CO 2.354 μm 2v
CH4 (group 1) 2.376 μm v1 + v4
CH4 (group 1) 2.425 μm v2 + 2v4

Figure 5 Pluto's spectrum and identification of various absorption bands. Reproduced from Doute S, Schmitt B, Quirico E, et al. (1999) Evidence for methane segregation at the surface of Pluto. *Icarus* 142: 421–444.

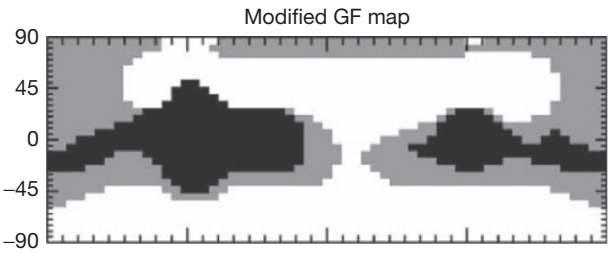


Figure 6 A Pluto composition map in cylindrical projection (Grundy and Buie, 2001). White areas are N_2-rich terrains, gray areas are CH_4-rich terrains, and black areas are H_2O/tholin-rich terrains.

adjacent to the eclipse event. The resulting visible-wavelength spectra showed Charon to be neutral in color, but the near-IR spectra of Charon obtained this way detected water ice on Charon's surface.

HST IR observations using the NICMOS IR imaging spectrometer (Buie and Grundy, 2000) were later obtained at four different longitudes in 1998. These spectra revealed that the near-IR spectrum of Charon varies only weakly with longitude and is consistent with crystalline water ice with grain sizes in the tens of micrometers (Brown and Calvin, 1999; Buie and Grundy, 2000), plus an unknown spectrally indistinct absorber. Ground-based IR spectra obtained in very good seeing conditions using the Keck 10-meter telescope (Brown and Calvin, 1999) also revealed evidence for crystalline H_2O ice and the possible presence of ammonia hydrates, suggesting the possibility of geologic activity. Cook et al. (2006) later confirmed these findings.

Spectroscopy of KBOs is considerably more difficult than Pluto and Charon because they are so much fainter. Most KBOs that have been studied show featureless reflectance spectra. However, strong 1.5 and 2.0 μm water ice features were

detected on the centaur (10199) Chariklo by Brown and Koresko (1998), on Pholus by Cruikshank et al. (1998), and on Chiron by Foster et al. (1999) and later Luu et al. (2000). Water ice features were first detected on a KBO by Brown et al. (1999) in a spectrum of 1996 TO66. Interestingly, these data indicated possible variations with rotational phase, perhaps indicating a patchy surface. Licandro et al. (2001) found deep water ice absorption in 2000 WR106. IR observations of the KBO (50000) Quaoar detected both crystalline water ice and ammonia hydrate (Jewitt and Luu, 2004). These data were interpreted to indicate that Quaoar has recently been resurfaced from either impacts or cryovolcanic activity. Crystalline water ice was also clearly detected in spectra of 2003 EL61 (Trujillo et al., 2006), again implying a resurfacing process taking place.

As to constituents other than water ice, the available data set is meager. Perhaps, the most interesting results are IR spectra of the large KBO 2005 FY9, which appears to be very similar to Pluto, but with even stronger prominent absorption bands of CH_4 (Licandro et al., 2006; see **Figure 7**). The tentative detection of N_2 on Sedna and Eris and CH_4 on Sedna has also been reported (Licandro et al., 2006).

10.20.3.5 Atmospheres

The only known KBO with an atmosphere is Pluto, which possesses an atmosphere with a base pressure of order 10 μb and a composition that is largely N_2, with traces of CO and CH_4. For excellent reviews, see Trafton et al. (1997) and Yelle and Elliot (1997).

Perhaps, the most important aspect of Pluto's atmosphere for the purpose of this review is that it seasonally collapses to the surface as Pluto draws away from the Sun and cools (e.g., Stern and Trafton, 1984; Trafton and Stern, 1983). This results in the seasonal/orbital transport and deposition of a high-albedo frost of thickness some millimeters to centimeters that persists for most of Pluto's orbit, when the atmosphere cannot be supported by thermally driven ice sublimation (Stern et al., 1988).

In order to support a significant atmosphere, a KBO must either exhibit internal outgassing or possess volatile ices that are warmed sufficiently to generate a nonnegligible vapor pressure equilibrium or both. Indeed, with the discovery of CH_4 frost on Pluto's surface (Cruikshank et al., 1976), the possibility of an atmosphere supported by vapor pressure equilibrium was already inferred.

This same process will occur on KBOs that generate seasonally or orbitally forced atmospheres, if the KBO is large enough to retain its atmosphere against immediate hydrodynamic (i.e., comet-like) escape. The estimated diameter for a KBO to retain its bound atmosphere is about 700 km (Lykawka and Mukai, 2005). For KBOs in excess of this size that possess atmospheres, one would expect a cyclic 'atmospheric laundry' to renew the surface albedo, as at Pluto (Stern et al., 1988).

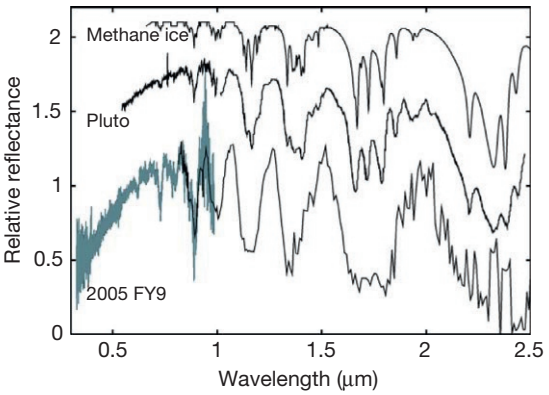


Figure 7 The spectrum of 2005 FY9 compared to Pluto and pure methane ice (Licandro et al., 2006); the latter two spectra were shifted vertically for clarity.

10.20.3.5.1 Pluto's atmosphere

Since vapor pressures have an exponential dependence in temperature, frost temperature is a critical parameter for understanding the density and pressure of Pluto's atmosphere. Temperature observations of Pluto's surface first measured the flux at the IR and millimeter wavelengths. Altenhoff et al. (1988), Stern et al. (1993), and Jewitt (1994) found brightness temperatures of ~30–40 K from millimeter measurements at 800–1200 mm. A similar result was obtained using the temperature-diagnostic 2.148 mm feature of N_2 as a thermometer (Tryka et al., 1994). However, Infrared Astronomical Satellite (IRAS) measurements by Sykes et al. (1987) found the brightness temperature to be warmer (55 K) at 60 mm, near the peak of Pluto's blackbody emission. The discrepancy between the *IRAS* and larger wave temperature measurements was and still is attributed to the presence of both hot (inert) and cold (sublimating) regions on Pluto's surface. Stern et al. (1993) predicted that if the latter was in fact the case, Pluto's visible and thermal light curves should be inversely correlated. Subsequently, observations using the ISO spacecraft detected a thermal light curve for the Pluto–Charon system at 60 mm, indicating that Pluto's surface is not isothermal (Lellouch et al., 2000), with maximum dayside temperatures of 54–63 K.

Combining information from surface temperature measurements and surface spectroscopy, one infers that either N_2 or CO must dominate Pluto's atmosphere and that CH_4 must be present only in trace amounts. N_2 is expected to dominate CO based on Raoult's law, that is, because it is dominant on the surface and the two have similar volatilities.

It was not until 1988 that Pluto's atmosphere was definitively detected using the technique of stellar occultation (Elliot et al., 1989; Hubbard et al., 1988). All of the 1988 stellar occultation light curves exhibit a sharp change of slope, or 'knee' during immersion and emersion (see **Figure 8**).

The structure of Pluto's atmosphere as probed by the 1988 occultation was well modeled by an isothermal upper atmosphere (above a radius of 1215 km) with a temperature of 104 ± 2 K for composition of nitrogen. This temperature is significantly greater than any surface temperature measurements for Pluto. The 'knee' in the occultation light curve can be explained by either a temperature gradient of 10–30 K km^{-1} (Eshelman, 1989; Hubbard et al., 1990; Stansberry et al., 1994) or a haze layer (Elliot et al., 1989). These mechanisms are not mutually exclusive and both may be or have then been at work.

Yelle and Lunine (1989) realized that CH_4 can act as a thermostat in Pluto's upper atmosphere by absorbing energy at the ν_3 band at 3.3 mm, cooling via the ν_4 band at 7.6 μm, and conducting heat to the surface. This mechanism may explain the high temperatures at the microbar pressure level in light of the colder surface temperatures. Strobel et al. (1996) expanded on this model and developed a radiative–conductive atmospheric model, including the effects of heating and cooling by methane at 3.3, 2.3, and 7.6 μm, cooling by CO rotational line emission (originally considered by Lellouch et al. (2000)), and heat conduction. Doing so, they found that the 2.3 μm band was more important for heating the atmosphere than the 3.3 μm band. Their preferred model can reproduce the temperature gradient of the occultation light curve but has an upper atmosphere temperature that is too warm (134 K).

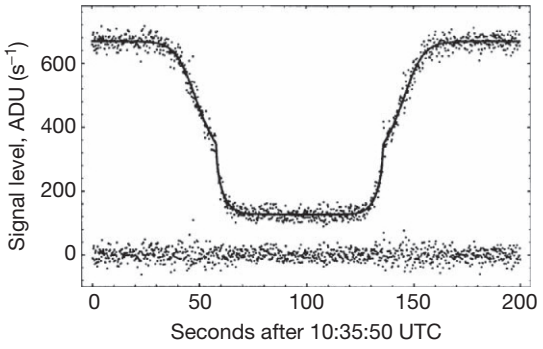


Figure 8 Data (points; Elliot et al., 1989) and model (line; Elliot and Young, 1992) of the 1988 stellar occultation light curve of Pluto observed with NASA's Kuiper Airborne Observatory. The points near the zero level along the bottom are the residuals from the fit of the model to the data.

Of course, all of the 'CH_4-thermostat' models for Pluto's atmospheric thermal structure rely on sufficient quantities of CH_4 in the atmosphere. And while CH4 had been inferred indirectly from the combination of its known presence in solid state on the surface and surface thermal measurements, its actual atmospheric abundance was highly uncertain because it was not known if the CH_4 source regions were in Pluto's cold or hot regions.

Spectroscopic detection of gaseous CH_4 was first obtained in 1994 (Young et al., 1997) with a subsequent confirmation in 2006 (Roe, 2006). This work confirmed that the gaseous CH_4 in Pluto's atmosphere is only a trace constituent. The partial pressure of methane is overabundant for methane in solid solution with nitrogen but consistent with methane in vapor pressure equilibrium with pure CH_4 at 41–45 K, as shown in **Figure 9**.

From the well-observed stellar occultation of August 2002, it was discovered that Pluto's atmospheric pressure had doubled since 1988 (Elliot et al., 2003; Sicardy et al., 2003; **Figure 10**). The increase in atmospheric pressure is consistent with a change in nitrogen frost temperature by 1.3 K (Elliot et al., 2003), which may be due to a thermal phase lag following the 1989 perihelion maximum in insolation (e.g., Stern et al., 1988; Trafton, 1990; see also the review by Spencer et al. (1997)).

In addition to the increase in atmospheric pressure, an increase of turbulence in Pluto's atmosphere was evidenced by more refractive spikes in the occultation light curves of 2002, compared with 1988. These spikes indicate the presence of waves or turbulence and were practically nonexistent in the 1988 observations. These spikes were also seen in the lower atmosphere in the 2002 observations (Pasachoff et al., 2005).

Another important change seen from 1988 to 2002 is the declining strength of the kink in the occultation light curves,

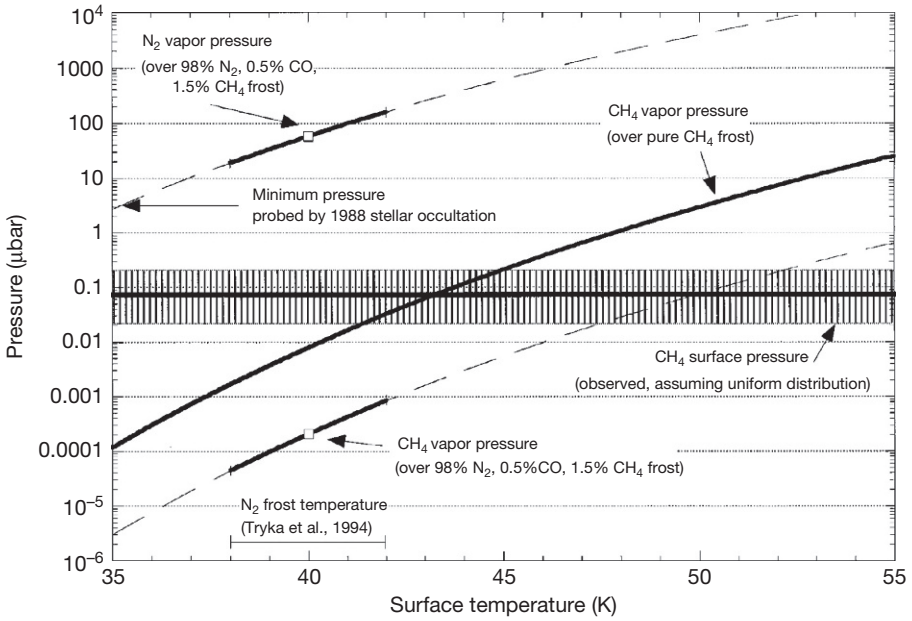


Figure 9 Surface vapor pressures on Pluto as a function of temperature of the sublimating material. The solid parts of the various curves show the vapor pressures for a frost temperature of 40 ± 2 K. Reproduced from Young LA, Elliot JL, Tokunaga A, de Bergh C, and Owen T (1997) Detection of gaseous methane on Pluto. *Icarus* 127: 258–270.

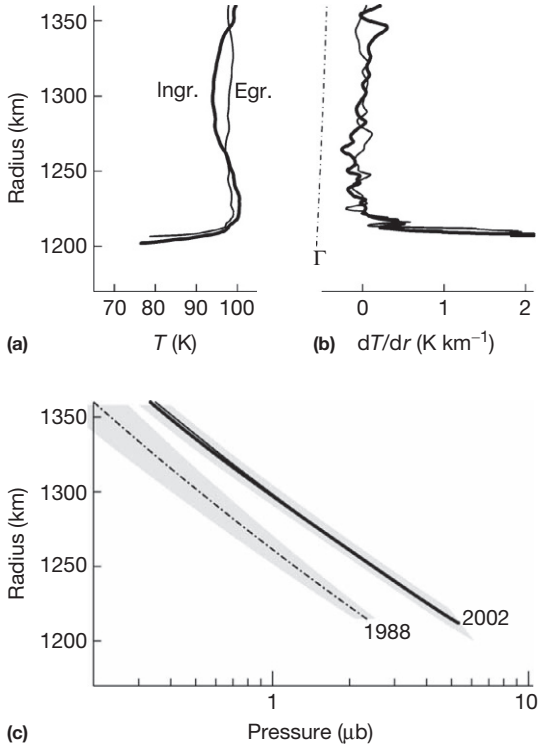


Figure 10 Temperature and pressure profiles of Pluto's atmosphere derived by Sicardy et al. (2003) from the inversion of the occultation light curve, assuming a spherically symmetrical and transparent pure nitrogen atmosphere. (a) Pluto's atmospheric temperature profiles $T(r)$ at ingress (thick curve) and egress (thin curve). The radius is the distance to the planet center. (b) The corresponding temperature gradient profiles dT/dr. The dash-dotted line provides the adiabatic lapse rate profile $\Gamma = -g/c_p$, to the left of which the atmosphere would become convectively unstable, where $c_p = 103\ \mathrm{J\ K^{-1}\ kg^{-3}}$ is the specific heat at constant pressure for N_2. (c) The corresponding pressure profiles. The dash-dotted line is the pressure profile derived from the Kuiper Airborne Observatory light curve of 9 June 1988, with the associated error domain as the shaded region. The ingress and egress profiles derived from the Canada–France–Hawaii Telescope (CFHT) light curve of 2002 are the nearly coincident thick and thin solid lines, respectively.

relative to what was seen in 1988. This indicates that a change in the atmosphere's vertical structure has also taken place. Elliot et al. (2003) reported on observations of the August 2002 occultation at effective wavelengths from 0.75 to 2.2 μm. The minimum flux of the light curve varies with wavelength, indicating that there is extinction in Pluto's lower atmosphere, that is, a haze layer. A purely refracting atmosphere would produce a light curve with the same minimum flux regardless of the wavelength.

All of the effects seen in the 2002 occultations were seen again in 2006, but the number of turbulence-driven refractory spikes appears to have further increased (e.g., Elliot et al., 2006; Sicardy et al., 2006; Young et al., 2006). Clearly, Pluto's atmosphere is changing, most likely due to postperihelion changes in insolation and subsolar latitude. When will atmospheric collapse take place? We do not know, but models (e.g., Trafton, 1990; Trafton et al., 1997) indicate it will occur before 2020.

10.20.3.5.2 Other KBOs

Regarding centaurs, 2060 Chiron has long been known to exhibit both photometric variability (Hartmann et al., 1989) and a cometary coma (Meech and Belton, 1989). Somewhat surprisingly, Chiron has exhibited activity at all points of its orbit with no clear correlation between the level of cometary activity and heliocentric distance. In fact, Chiron was more active at its aphelion in 1970 than it was in following decades when it was closer to the Sun. The level of activity can vary by more than a factor of 2 and last anywhere from weeks to months. Bus et al. (1991) detected CN in Chiron's coma, and Womack and Stern (1999) reported a detection of CO, which may be the sublimation agent responsible for Chiron's cometary activity. This suggestion is consistent with the models indicating that Chiron's activity can be explained by a composition of CO near the surface as an ice or as a gas trapped in the amorphous ice that is released by delayed thermal pulses (Capria et al., 2000; Prialnik et al., 1995). The variability of Chiron's outbursts, the apparent lack of a correlation between activity and distance from the Sun, and the observations of discrete jetlike features observed during an occultation (Elliot et al., 1995) imply that Chiron's surface may be unevenly covered with volatile material and may exhibit geysers like those seen on Triton (e.g., Smith et al., 1989).

In contrast to the known atmosphere around Pluto, only upper limits have been placed on possible atmospheres around other KB bodies.

Although no atmospheres have as yet (November 2006) been detected around KBOs other than Pluto, Hainaut et al. (2000) found a change in the light curve of KBO 1996 TO66 from a double peak to a single peak over the course of a year, which they suggested may have resulted from an episode of cometary activity. Further, the discoveries of N_2, CO, CH_4, and C_2H_6 (ethane) on the KBO 2005 FY9 (Brown et al., 2006; Licandro et al., 2006), and the tentative detection of N_2 on Sedna and Eris and CH_4 on Sedna (Licandro et al., 2006), indicate that KBO atmospheres may be discovered in the future. However, given the exponentially sensitive dependence of vapor pressures to temperature and therefore to heliocentric distance and seasonal (obliquity-driven) effects, it is natural to predict that such atmospheres should in many cases be transient with seasonally or orbitally driven cycles, as Pluto's is.

Finally, it is worth noting that the captured dwarf planet Triton, now a satellite of Neptune, is known to possess an N_2-dominated atmosphere with some CH_4 and a base pressure not unlike Pluto's. Since Triton is also roughly Pluto's size and density and is widely thought to have originated in the same region of the solar system as Pluto, it serves as a possible guide to future discoveries that may be made around dwarf planets in the KB and farther out as well.

10.20.3.6 Densities

With the discovery of two new satellites in the Pluto system (Weaver et al., 2006), Buie et al. (2006) had been able to derive a better estimate of the mass ratio of Charon to Pluto, and

therefore the individual masses of each, than had previously been possible. Combining this new mass ratio with the recently improved 604 ± 3 km radius of Charon from the stellar occultation in 2005 (Sicardy et al., 2006), we now also have a much more accurate estimate for Charon's density, 1.66 ± 0.06 g cm^{-3} (Buie et al., 2006). The current best estimate of Pluto's density is 2.03 ± 0.06 g cm^{-3} (Buie et al., 2006).

Densities for KBOs are only available for a subset of the two dozen objects with satellites. In such binary systems, we can determine the mass ratio of the satellite to the primary from the KBO satellite mutual orbit, but we also need an estimate of the diameter to derive the density; the latter requires direct resolution of the KBO's size by HST, thermal IR measurements, or mutual events – all of which are difficult to obtain. As a result, only a few KBO and centaur densities are available. Spencer et al. (2006) reported a density of 0.51(+0.29, −0.11) g cm^{-3} for KBO (26308) 1998 SM165; this is comparable to the 0.5–0.8 g cm^{-3} density derived for KBO (47171) 1999 TC36 (Stansberry et al., 2006). Grundy et al. (2007) reported a density for centaur (65489) 2003 FX128 of 1.38(+0.65, −0.32) g cm^{-3}. However, Rabinowitz et al. (2006) reported a far higher, model-dependent density of 2.6–3.3 g cm^{-3} for 2003 EL61 from studies of this KBO's rotational dynamics.

Clearly, the study of KBO densities is in its infancy. Only a handful of results are available. All concern large KBOs and all but one concern binary KBOs. However, already there appears to be a wide diversity in KBO densities, ranging from ~0.5 to >2 g cm^{-3}. This in turn implies a wide range of internal porosities and water ice/rock fractions in the KBO population.

10.20.3.7 Interior Models

Several groups have begun to investigate the thermal evolution of KBOs. One motivation for such studies is the current-day presence of surface volatiles on Pluto after what appears to be 4 Gy of active atmospheric escape. As demonstrated some time ago (Stern, 1989a), this argues for the possibility of present-day internal activity to pipe new volatiles to compensate escape losses. Another motivation for such studies results from the sporadic activity of the 160 km wide centaur Chiron, which shows no correlation with heliocentric distance, solar activity, or obliquity-derived season (e.g., Bus et al., 1991; Duffard et al., 2002).

Models show that radiogenic heating and solar irradiation of an initially undifferentiated KBO can sublime and deplete most of the volatile ices near the surface. The resulting internal heating can result in a compositionally layered structure, with interlaced layers depleted and enriched in volatiles. When such bodies are injected into the inner solar system, episodic activity can occur as the volatile-rich layers reach sublimation temperatures. Thermal evolution can also be affected by the heating of collisions, which can sublimate ices, induce gas flow in porous structures, and eject gas and dust. Depths thermally affected by KBO impacts can vary by large amounts and to depths greater than 1 km (Orosei et al., 2001).

To better understand interior evolution, McKinnon (2002) and Choi et al. (2002) conducted studies of the thermal evolution of KBOs of various sizes up to almost 1000 km in diameter. These workers agree that ^{26}Al cannot be an important heat source, owing to the long accretion timescale for KBOs relative to the 0.73 My half-life of ^{26}Al. They also found that for 100 km KBOs, radiative losses to space prevent central temperatures from exceeding 80 K; for 150 km KBOs, the central temperature limit found was near 105 K. However, for a 450 km radius KBO, central temperatures can exceed 270 K; this opens the possibility of liquid H_2O in large KBOs and Pluto.

Moreover, McKinnon (2002) found that even KBOs in the 75–225 km radius range can transform H_2O ice from its amorphous to crystalline state in their interiors. Other important thermal benchmarks include 155 K, where a eutectic CH_3OH–H_2O melt can form, and 190 K, where solid-state H_2O ice creep and hence sintering/densification become important. McKinnon (2002) concluded that over a wide range of parameter choices, long-lived radioactive energy release probably powered activity in the largest KBOs for hundreds of millions of years, at least. These findings indicate Pluto and other large KBOs may well show early, and perhaps even persistent, signs of internal activity.

Most recently, McKinnon and Barr (2006) and Hussman et al. (2006) had discussed the possibility of internal liquid H_2O oceans in large KBOs, including both Pluto and Charon. For rock ratios of 0.65 (consistent with Pluto's density) and surface temperatures near 40 K, internal oceans were commonly found, though at a range of interior depths and with various thicknesses. This exciting possibility begs for tighter constraints on KBO densities and compositions, so as to better constrain the range of parameter space available to modelers.

10.20.4 Origin

Unlike early ideas in which the trans-Neptunian debris belt contains only leftover material in the form of very small bodies from the era of giant planet formation (e.g., Kuiper, 1951), the real KB actually contains very large numbers (i.e., $>10^5$) of bodies that reached sizes of 100 km and at least several dozen and more likely many times more objects (depending on the large body population of the SKB) that reached 1000–3000 km in diameter.

The standard paradigm of planet formation involves three major stages. We review this paradigm briefly here but refer the reader to excellent, succinct reviews with more detail, such as Ward (1996) and **Chapter 10.17**.

The first stage of planet formation involves the coagulation of dust grains into macroscopic objects that grow by gentle accretion when physical contact occurs but at a rate dominated purely by two-body, physical (i.e., projected surface area) collision cross sections. Stage 2, the so-called midstage of accretion, begins when collision rates are substantially affected by gravitational cross sections, which accelerates growth from an R^2 process to an R^4 process. Midstage growth terminates when bodies self-isolate by growing to a size where they have depleted the feedstock of smaller material within their gravitational reach. The third and final stage of accretion results from the mergers of the planetary 'embryos' created in stage 2. These mergers occur on longer timescales driven by dynamic effects that cause the embryos to perturb one another onto mutually

crossing orbits that inevitably cause collisions among relatively like-sized bodies.

In order for solid-body accretion to proceed, two key conditions must be simultaneously obtained on a relevant timescale. These are (1) a sufficient supply of material to promote collisions on the relevant timescale and (2) sufficiently gentle collisions to permit growth, rather than erosion and/or destruction.

Shortly after the discovery of the first KBOs, when crude estimates of the population of the KB first became available, it was recognized (Stern, 1995) that the current-day KB does not satisfy either of these two conditions for accretion. More specifically, it was pointed out that (1) the surface mass density of the current-day KB is too low by a factor of approximately 10 to allow the growth of the observed KBOs in the age of the solar system and (2) current-day collision velocities (calculated from the mean random crossing speeds of intersecting orbits) are too high by a factor of about 10^4 to initiate and sustain stage 1 growth. These facts in turn led to the conclusion (Stern, 1995, 1996) that the current KB must be highly evolved from its primordial state, in which both the mass density of material must have been far higher and the mean random orbital inclinations and eccentricities must have been much lower.

Subsequent, more detailed accretional simulations (e.g., Davis and Farinella, 1997; Kenyon and Bromley, 2001; Kenyon and Luu, 1998, 1999; Stern and Colwell, 1997a,b) reached similar conclusions and refined an accretion scenario for the formation of the KB that can be summarized as follows. The early KB region between 30 and 50 AU originally contained some 10–50 M_{Earth} of material in the form of dust and small debris. The region was dynamically cold, so that collisions promoted growth. It was likely located somewhat closer to the Sun, perhaps by as much as 5–10 AU, as testified to by Neptune's now well-established migration (e.g., Malhotra, 1993). After order 100–300 My, objects of size 30–50 km had emerged on nearly circular orbits. These objects were of sufficient size to accelerate their own growth by gravitational runaway. During this midstage accretion, objects of size 100–2000 km were rapidly grown. At this point, dynamic excitation of the region was terminated, either by the nearby growth of Neptune or by Mars-sized objects either grown or scattered into the region. This terminated growth by transforming the accretional environment to an erosional one, with characteristic collision velocities of order 1–2 km s^{-1}, as are observed in the KB today. The onset of erosion initiated a fratricide among small bodies, which resulted in the loss of much mass to grinding followed by radiation transport out of the region (Stern and Colwell, 1997b). (Bodies hundreds of kilometers in diameter and larger, by virtue of their mass and gravitational binding energy, were relatively immune to this fratricide.) It was estimated that over 4 Gy, erosional mass loss could deplete the KB down to a few tenths of an Earth mass of material. Although this is higher than the observed KB mass by a factor of a few to perhaps 30 (in the extreme), dynamically driven losses have been shown to be capable of further mass depletion by a factor of 2 to several, thereby making this scenario viable. In the in situ accretion scenario (ISAS), plutinos, which migrated to their current orbits after being trapped in Neptune MMRs during their outward migration, preferentially formed at somewhat closer distances (e.g., 20–30 AU) than nonresonant CKBOs and SKBOs.

The salient features of this ISAS are that (1) the early KB region initially contained an amount of mass consistent with a simple extrapolation of the surface mass density of the giant planet region out to the 20–50 AU zone and (2) growth proceeded until excitation by external (or perhaps internal) perturbers interceded, some time between 0.1 and 1 Gy after the formation of the protosolar disk.

An alternative to the ISAS described just earlier has recently developed (e.g., Morbidelli, 2006). This scenario, called the 'push out scenario' (POS), was motivated by dynamicists attempting to recreate various aspects of the orbital distribution of KBOs. In the POS, all of the objects in the KB were accreted in the giant planets' region (5–25 AU), rather than beyond this region as the ISAS assumes. These objects were then subsequently transported outward by a combination of gravitational clearing (i.e., scattering) by the giant planets and resonance migration driven by the outward migration of Neptune to conserve angular momentum during the clearing of the giant planets' zone.

The ISAS and POS each have strengths and weaknesses, but a detailed discussion of such is beyond the scope of this chapter. So we restrict our discussion to the point that while the POS points to a compact solar nebula, initially truncated inside ~20 AU, the ISAS points to a broader, dynamically cold disk, stretching out to distances of ~50 AU or even much farther.

Therefore, a distinguishing test between these scenarios surrounds the issue of whether there is a dynamically cold population of KBOs or smaller debris lying farther from the Sun than the ~50 AU cutoff of the cold KBO population now observed. Existing observational constraints make clear there is no such cold disk of large KBOs between ~50 and ~90 AU. However, this could be the result of a dynamic clearing of this region due to the former presence of a passing star or large planets temporarily scattered there during outer solar system clearing or due to a change in the accretional environment beyond 50 AU that stunted the growth of bodies there at much smaller sizes than the KBOs in the 30–50 AU zone reached (making such bodies less detectable). Future deep surveys will be required to resolve the question of whether any substantial cold disk lies beyond ~50 AU. Such work must constrain both the population density of small (e.g., 1–10 km) bodies on low-eccentricity, low-inclination orbits in the 50–100 AU zone and the population of large (e.g., >100 km) bodies on low-eccentricity, low-inclination orbits in the region beyond 100 AU.

10.20.5 Future Observational Goals and Prospects

Existing KBO data and models provide rich insights into this region of the solar system, but, realistically, this field of planetary research is still in its infancy. In the next couple of decades, we expect to see far faster growth in our knowledge and understanding of the KB than in the past, driven by new ground-based observatories, space-based telescopes, space missions (see **Chapter 10.23**), and computer and detector advances. In this section, we examine some of the highest priority objectives for future KB studies and then discuss the expected advances in ground-based and space-based observations that are expected in the next decade.

10.20.5.1 Goals

The broad goals for the next stage of observational study of KBOs are relatively straightforward. They are to (1) develop a broad understanding of the various populations in the KB and (2) determine the evolutionary processes that have shaped KBOs to their current state. The first of these goals requires collecting large surveys of albedo, color, composition, shape, and density information. It also requires thorough searches for KBO atmospheres and the determination of KBO satellite size–frequency–orbital types. It is also strongly desired to begin studying surface heterogeneity on KBOs other than Pluto; this includes albedo, color, and compositional heterogeneity. An additional goal should be to better define the KBO size–frequency histograms down to sizes of kilometers for each orbit type in the KB.

Of particular interest, we point out that binary and higher-order multiple systems offer great potential as tools for studying the physical properties of their components. Knowledge of the orbit of the secondary relative to the primary provides a system mass, and component masses can be determined if absolute astrometry is available to provide motions for each component. Relative sizes can be determined from photometry if one assumes equal reflectance, and absolute sizes inferred if one assumes particular values for the albedos. From sizes and masses, densities can be determined, providing information regarding interior properties and constituents. Of particular interest will be mutual events, like those observed between Pluto and Charon from 1985 to 1990. Such events occur when the line of sight from the Earth intersects the orbital plane of the satellite, so that the parent and satellite(s) occult and transit each other. Photometric and spectroscopic observations during these occultations can provide measurements of absolute sizes, probe atmospheres, and identify surface features and properties on scales far smaller than the resolution limit of the telescopes.

10.20.5.2 Future Ground-Based Observatories

Several large-scale survey telescopes with revolutionary capabilities for KB studies are currently in development, notably the Large Synoptic Survey Telescope (LSST) and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS). These projects are designed to observe the entire sky observable from the site at a daily or weekly frequency.

The significant new aspect of both these projects compared to previous surveys is to provide regular monitoring of 'variation' of objects over the full sky. These projects will provide revolutionary new capabilities for surveying the content of the KB, obtaining large numbers of accurate orbits, and studying KBO light curves, phase curves, colors, and satellite orbits.

The LSST (e.g., Davison and Angel, 2002; Tyson, 2002) is an 8.4 m telescope that will survey the entire available sky every three nights with multiple color filters. Its 3.5-degree CCD field of view will generate 3 billion pixels per image with a plate scale that will be close to 0.2 arcsec/pixel. Some 30 terabytes of data will be produced per night, requiring massive storage and a fast data-processing pipeline. LSST's performance specifications mentioned earlier indicate it will accomplish a complete survey of KBOs brighter than 25th magnitude every three nights, which will find ~50 000 KBOs on the order of 100 km in diameter. Potential observatory sites have been narrowed to three: Cerro Pachon (in Chile), Las Campanas (in Chile), and San Pedro Mártir (in Mexico). The project goal is to begin operations in 2012.

Pan-STARRS (Hodapp et al., 2004; Jewitt, 2003; Kaiser et al., 2002) will take a different approach to a similar goal. It will consist of four 1.8 m telescopes coordinated to observe the same part of the sky simultaneously with a roughly 3 square degree field of view and a plate scale of 0.3 arcsec. Exposure times of 30–60 s are expected to reach a limiting magnitude of 24. The sky from its site in Hawaii will be observed three times during the darkest nights of each lunation. The first prototype telescope for Pan-STARRS had first light in early 2006, and the full system is planned to be in operation before 2010.

After 3–5 years of operations, it is expected that these projects will produce a complete, flux-limited sample of KBOs that should be free of observational biases. Many tens of thousands of new KBOs should be discovered. This will yield a census of the populations of all known and any major as-yet-undiscovered dynamic groups, as well as substructures therein. They will also provide a vast database of KBO light curves, phase curves, colors, and other data, as noted earlier.

Smaller-scale 'pencil beam' and wide-area surveys to even greater depth using large telescopes will continue to be necessary to probe the extremes of the size distribution of objects in the KB.

Another approach will be taken by TAOS, the Taiwan–America Occultation Survey (Alcock et al., 2003). The goal of TAOS is to sample the low end of the size distribution, by detecting occultations of stars by KBOs with sizes down to a few kilometers. TAOS is designed to compensate for the weakness of doing direct detection of KBOs by reflected light, in which the brightness of the objects decreases approximately by its distance to the fourth power owing to the compound inverse square law of reflected light sources. TAOS will employ three small (20 in.) robotic telescopes with 2-degree fields of view. They monitor a fixed set of stars throughout the year, observing the same areas of sky simultaneously, and possible events are cross compared to reject false detections. The third TAOS telescope was installed in April 2004, but the system is still in test as of late 2006. Although detections of asteroid occultations have been made, no definitive KBO occultation events have been detected.

Other ground-based advances to look forward to in the coming 10–15 years include

- occultation campaigns on known KBOs to better determine their sizes and albedos,
- photometric and spectroscopic campaigns to study mutual event seasons for KBOs with satellites, and
- spectroscopy of KBOs using a new generation of very large optical telescopes coming on line near the middle to end of the 2010–20 decade.

10.20.5.3 Future Space-Based Observations

Two new major space mission capabilities are now under development. We discuss both here. We do not describe missions in study phase or proposals for future missions.

The HST will be able to continue obtaining observations for several more years. If a space shuttle servicing mission occurs as planned in 2008, it will install new gyros, batteries, and a fine guidance sensor, with two new instruments – the Wide Field Camera 3 (WFC3) and the Cosmic Origins Spectrograph (COS). WFC3 will have greater sensitivity than WFC2, allowing for improved photometry and imaging of fainter objects. COS will restore HST's UV spectroscopy capabilities and extend them to sources a factor of about 20 times fainter than other HST spectrographs allowed.

The James Webb Space Telescope (JWST), formerly the Next Generation Space Telescope (NGST), will be a large, IR-optimized space telescope to follow HST. JWST is expected to be launched to the second Lagrange point (L2) in 2013–15 and has an expected lifetime of 5–10 years. It will have a 6.5 m diameter mirror (made of 18 hexagonal segments) and an IR camera and spectrograph, providing higher resolution and perhaps a couple or more orders of magnitude greater sensitivity than HST. Studies by the NGST working group estimate that 10 km diameter objects could be detected at distances of 50 AU. However, JWST will not be able to observe at long enough wavelengths (for the thermal peak around 75 mm) to determine definitive albedos and sizes of KBOs, as can be done with the Spitzer Space Telescope (Emery et al., 2004; Grundy et al., 2005a,b; Stansberry et al., 2004). JWST will be able to study the IR spectra of numerous KBOs and their satellites.

10.20.5.4 Space Missions to the KB and Centaurs

NASA's *New Horizons* mission (Stern and Cheng, 2002) to Pluto and the KB successfully launched on 19 January 2006. After a gravity assist flyby of Jupiter in March 2007, the spacecraft will make its closest approach flyby of the Pluto system on 14 July 2015. *New Horizons* will then go on to conduct one or possibly two KBO encounters in the succeeding 5–6 years. After that time, the spacecraft will emerge from the main region of the KB and into the depleted region beyond 50 AU. KBO targets will be chosen in 2014.

New Horizons carries a suite of seven instruments that include visible light images, IR mapping spectrometers, UV mapping spectrometers, radio science, radiometry, in situ plasma detectors, and dust detectors.

Its primary objectives at each flyby target are to map surface compositions and temperatures, characterize geology and morphology, characterize (or search for) the neutral atmospheres, and study escape rates. Secondary objectives include searches for new satellites and rings, ionospheric studies, stereo mapping, and refinement of bulk parameters such as masses, densities, and orbits.

No missions to centaur bodies are in development yet, but *New Horizons* will conduct a distant (~2.8 AU) flyby of centaur 2002 GO9 in 2009–10, allowing the first-ever measurement of a phase curve for studying the photometric and Bond albedo of such a body.

No additional missions to bodies in the KB are currently in development, but the US Decadal Survey of Planetary Science did place high priority on exploring the diversity of KBOs with space missions. Hence, it would not be surprising to see proposals for future missions.

References

Alcock C, Dave R, Giammarco J, et al. (2003) TAOS: The Taiwanese–American Occultation Survey. *Earth Moon and Planets* 92: 459–464.

Altenhoff WJ, Chini R, Hein H, et al. (1988) First radio astronomical estimate of the temperature of Pluto. *Astronomy and Astrophysics* 190: L15–L17.

Barkume KM, Brown ME, and Schalller EL (2006) Water ice on the satellite of Kuiper Belt object 2003 EL61. *Astrophysical Journal* 640: L87–L89.

Barucci MA, de Bergh C, Cuby J-G, Le Bras A, Schmitt B, and Romon J (2000a) Infrared spectroscopy of the Centaur 8405 Asbolus: First observations at ESO-VLT. *Astronomy and Astrophysics* 357: L53–L56.

Barucci MA, Doressoundiram A, Tholen D, Fulchignoni M, and Lazzarin M (1999a) Spectrophotometric observations of Edgeworth–Kuiper Belt objects. *Icarus* 142: 476–481.

Barucci MA, Fulchignoni M, Birlan M, Doressoundiram A, Romon J, and Boehnhardt H (2001) Analysis of trans-Neptunian and Centaur colours: Continuous trend or grouping? *Astronomy and Astrophysics* 371: 1150–1154.

Barucci MA, Lazzarin M, and Tozzi GP (1999b) Compositional surface variety among the Centaurs. *Astronomical Journal* 117: 1929–1932.

Barucci MA, Romon J, Doressoundiram A, and Tholen DJ (2000b) Compositional surface diversity in the trans-Neptunian objects. *Astronomical Journal* 120: 496–500.

Bernstein GM, Trilling DE, Allen RL, Brown ME, Holman M, and Malhotra R (2004) The size distribution of trans-Neptunian bodies. *Astronomical Journal* 128: 1364–1390.

Bertoldi F, Altenhoff W, Weiss A, Menten KM, and Thum C (2006) The trans-Neptunian object UB313 is larger than Pluto. *Nature* 439: 563.

Binzel RP (1988) Hemispherical color differences on Pluto and Charon. *Science* 241: 1070–1072.

Binzel RP (1989) Pluto–Charon mutual events. *Geophysical Research Letters* 16: 1205–1208.

Binzel RP, Tholen DJ, Tedesco EF, Buratti BJ, and Nelson RM (1985) The detection of eclipses in the Pluto–Charon system. *Science* 228: 1193–1195.

Boehnhardt H, Tozzi GP, Birkle K, et al. (2001) Visible and near-IR observations of trans-Neptunian objects. Results from ESO and Calar Alto Telescopes. *Astronomy and Astrophysics* 378: 653–667.

Brown ME (2001) The inclination distribution of the Kuiper belt. *Astronomical Journal* 121: 2804–2814.

Brown ME and Calvin WM (1999) Evidence for crystalline water and ammonia ices on Pluto's satellite Charon. *Science* 287: 107–109.

Brown ME and Koresko CC (1998) Detection of water ice on the centaur 1997 CU 26. *Astrophysical Journal Letters* 505: L65–L67.

Brown ME, Van Dam M, Bouchez AH, et al. (2006) Satellites of the largest Kuiper Belt objects. *Astrophysical Journal* 639: L43–L46.

Brown RH, Cruikshank DP, and Pendleton Y (1999) Water ice on Kuiper Belt object 1996 TO66. *Astrophysical Journal* 519: L101–L104.

Buie MW, Cruikshank DP, Lebofsky LA, and Tedesco EF (1987) Water frost on Charon. *Nature* 329: 522–523.

Buie MW and Grundy WM (2000) The distribution and physical state of H_2O on Charon. *Icarus* 148: 324–339.

Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2006) Orbits and photometry of Pluto's satellites: Charon, S/2005 P1 and S/2005 P2. *Astronomical Journal* 132: 290–298.

Buie MW, Tholen DJ, and Wasserman LH (1997) Separate light curves of Pluto and Charon. *Icarus* 125: 233–244.

Bus SJ, A'Hearn MF, Schleicher DG, and Bowell E (1991) Detection of CN emission from (2060) Chiron. *Science* 251: 774–777.

Capria MT, Coradini A, De Sanctis MC, and Orosei R (2000) Chiron activity and thermal evolution. *Astronomical Journal* 119: 3112–3118.

Choi Y, Cohen M, Merk R, and Prialnik D (2002) Long-term evolution of objects in the Kuiper Belt Zone – Effects of insolation and radiogenic heating. *Icarus* 160: 300–312.

Christy JW and Harrington RS (1978) The satellite of Pluto. *Astronomical Journal* 83: 1005.

Cohen CJ and Hubbard EC (1965) Libration of the close approaches of Pluto to Neptune. *Astronomical Journal* 70: 10.

Cook JC, Desch SJ, Roush TL, Geballe TR, and Trujullo CA (2006) Near-infrared spectroscopy of Charon: Possible evidence for cryovolcanism on Kuiper Belt objects. *AAS/DPS Meeting*, 21.02.

Cruikshank DP, Mason RE, Dalle Ore CM, et al. (2006) Ethane on Pluto and Triton. *AAS/DPS Meeting*, 21.03.

Cruikshank DP, Pilcher CB, and Morrison D (1976) Pluto – Evidence for methane frost. *Science* 194: 835–837.

Cruikshank DP, Roush TL, Bartholomew MJ, et al. (1998) The composition of Centaur 5145 Pholus. *Icarus* 135: 389–407.
Cruikshank DP, Stansberry JA, Emery JP, et al. (2005) The high-albedo Kuiper Belt object (55565) 2002 AW197. *Astrophysical Journal* 624: L53–L56.
Davies J (2001) *Beyond Pluto Exploring the Outer Limits of the Solar System*. Cambridge: Cambridge University Press.
Davis DR and Farinella P (1997) Collisional evolution of Edgeworth–Kuiper Belt objects. *Icarus* 125: 50–60.
Davison WB and Angel JRP (2002) Large synoptic survey telescope mechanical structure and design. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries Proceedings of the SPIE*, 4836, pp. 104–110.
Delsanti A and Jewitt D (2006) The solar system beyond the planets. In: Blondel Ph and Mason J (eds.) *Solar System Update*. Chichester: Praxis-Springer.
Delsanti AC, Boehnhardt H, Barrera L, Meech KJ, Sekiguchi T, and Hainaut OR (2001) BVRI photometry of 27 Kuiper Belt objects with ESO/very large telescope. *Astronomy and Astrophysics* 380: 347–358.
Dobrovolskis AR, Peale SJ, and Harris AW (1997) Dynamics of the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, p. 159. Tucson, AZ: University of Arizona Press.
Doressoundiram A, Barucci MA, Tozzi GP, et al. (2005) Spectral characteristics and modeling of the trans-Neptunian object (55565) 2002 AW197 and the Centaurs (55576) 2002 GB10 and (83982) 2002 GO9: ESO Large Program on TNOs and Centaurs. *Planetary and Space Science* 53: 1501–1509.
Douté S, Schmitt B, Quirico E, et al. (1999) Evidence for methane segregation at the surface of Pluto. *Icarus* 142: 421–444.
Duffard R, Lazzaro D, Pinto S, et al. (2002) New activity of Charon: Results from 5 years of photometric monitoring. *Icarus* 160: 44–51.
Duncan M, Quinn T, and Tremaine S (1987) The formation and extent of the solar system comet cloud. *Astronomical Journal* 94: 1330–1338.
Duncan M, Quinn T, and Tremaine S (1988) The origin of short-period comets. *Astrophysical Journal* 328: L69–L73.
Durda DD and Stern SA (2000) Collision rates in the present-day Kuiper Belt and Centaur regions: Applications to surface activation and modification on comets, Kuiper Belt objects, Centaurs and Pluto–Charon. *Icarus* 145: 220–229.
Edgeworth KE (1943) The evolution of our planetary system. *Journal of the British Astronomical Association* 53: 181–188.
Edgeworth KE (1949) The origin and evolution of the solar system. *Monthly Notices of the Royal Astronomical Society* 109: 600–609.
Elliot JL, Ates A, Babcock BA, et al. (2003) The recent expansion of Pluto's atmosphere. *Nature* 424: 165–168.
Elliot JL, Dunham EW, Bosh AS, et al. (1989) Pluto's atmosphere. *Icarus* 77: 148–170.
Elliot JL, Olkin CB, Dunham EW, et al. (1995) Jet-like features near the nucleus of Chiron. *Nature* 373: 46.
Elliot JL, Person MJ, Gulbis AA, et al. (2006). The size of Pluto's atmosphere as revealed by the 2006 June 12 occultation. *AAS/DPS Meeting*, 21.02.
Elliot JL and Young LA (1992) Analysis of stellar occultation data for planetary atmospheres: I: Model fitting, with application to Pluto. *Astronomical Journal* 103: 991–1015.
Emery JP, Cruikshank DP, Van Cleve J, and Stansberry JA (2004) Spitzer/IRS observations of asteroids, Centaurs, and Kuiper Belt objects. *AGU Fall Meeting Abstracts*, A978.
Eshelman VR (1989) Pluto's atmosphere: Models based on refraction, inversion, and vapor–pressure equilibrium. *Icarus* 80: 439–443.
Farinella P, Davis DR, and Stern SA (2000) Formation and collisional evolution of the Edgeworth–Kuiper belt. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 1255–1280. Tucson, AZ: University of Arizona Press.
Fernandez JA and Ip WH (1984) Some dynamical aspects of the accretion of Uranus and Neptune – The exchange of orbital angular momentum with planetesimals. *Icarus* 58: 109–120.
Fernández YR, Jewitt DC, and Sheppard SS (2002) Thermal properties of Centaurs Asbolus and Chiron. *Astronomical Journal* 123: 1050–1055.
Fink U and DiSanti MA (1988) The separate spectra of Pluto and its satellite Charon. *Astronomical Journal* 95: 229–236.
Fix JD, Neff JS, and Kelsey LA (1970) Spectrophotometry of Pluto. *Astronomical Journal* 75: 895.
Foster MJ, Green SF, McBride N, and Davies JK (1999) Detection of water ice on 2060 Chiron. *Icarus* 141: 408–410.
Gerakines PA, Moore MH, and Hudson RL (2001) Energetic processing of laboratory ice analogs: UV photolysis versus ion bombardment. *Journal of Geophysical Research* 106: 33381–33386.
Gil-Hutton R (2002) Color diversity among Kuiper Belt objects: The collisional resurfacing model revisited. *Planetary and Space Science* 50: 57.
Greenberg JM (1982) In: Wilkening LL (ed.) *Comets*, pp. 131–163. Tucson, AZ: University of Arizona Press.
Grundy WM and Buie MW (2001) Distribution and evolution of CH_4, N_2, and CO Ices on Pluto's surface: 1995 to 1998. *Icarus* 153: 248–263.
Grundy WM and Fink U (1996) Synoptic CCD spectrophotometry of Pluto over the past 15 years. *Icarus* 124: 329–343.
Grundy WM, Noll KS, and Stephens DC (2005a) Diverse albedos of small trans-Neptunian objects. *Icarus* 176: 184–191.
Grundy WM, Spencer JA, Noll KS, et al. (2007) The orbit, mass, size, albedo, and density of (65489) 2003 FX128: A tidally evolved Centaur. *Icarus* 191(1): 286–297.
Grundy WM, Spencer JR, Stansberry JA, et al. (2005b) Spitzer/MIPS survey of classical Kuiper Belt objects and a Neptune Trojan. *AAS/Division for Planetary Sciences Meeting Abstracts, 37*.
Hainaut OR, Delahodde CE, Boehnhardt H, et al. (2000) Physical properties of TNO 1996 TO66. Light curves and possible cometary activity. *Astronomy and Astrophysics* 356: 1076–1088.
Hainaut OR and Delsanti AC (2002) Colors of minor bodies in the outer solar system. A statistical analysis. *Astronomy and Astrophysics* 389: 641–664.
Hartmann WK, Cruikshank DP, and Tholen DJ (1989) 2060 Chiron. *Icarus* 83: 1–15.
Hodapp KW, Kaiser N, Aussel H, et al. (2004) Design of the Pan-STARRS telescopes. *Astronomische Nachrichten* 325: 636–642.
Hubbard WB, Hunten DM, Dieters SW, Hill KM, and Watson RD (1988) Occultation evidence for an atmosphere on Pluto. *Nature* 336: 452–454.
Hubbard WB, Yelle RV, and Lunine JI (1990) Non-isothermal Pluto atmosphere models. *Icarus* 84: 1–11.
Hudson RL and Moore MH (2001) Radiation chemical alterations in solar system ices: An overview. *Journal of Geophysical Research* 106: 33275–33284.
Hussman H, Sohl F, and Spohn T (2006) Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.
Jewitt D (2003) Project Pan-STARRS and the outer solar system. *Earth Moon and Planets* 92: 465–476.
Jewitt D and Luu J (1992) 1992 QB1. *International Astronomical Union Circular* 5611.
Jewitt D and Luu J (1998) Optical-infrared spectral diversity in the Kuiper Belt. *Astronomical Journal* 115: 1667–1670.
Jewitt DC (1994) Heat from Pluto. *Astronomical Journal* 107: 372–378.
Jewitt DC (2002) From Kuiper Belt object to cometary nucleus: The missing ultra-red matter. *Astronomical Journal* 123: 1039–1049.
Jewitt DC and Luu J (2004) Crystalline water ice on the Kuiper Belt object (50000) Quaoar. *Nature* 432: 731–733.
Jewitt DC and Luu JX (2001) Colors and spectra of Kuiper Belt objects. *Astronomical Journal* 122: 2099–2114.
Kaiser N, Aussel H, Burke BE, et al. (2002) Pan-STARRS: A large synoptic survey telescope array. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries, Proceedings of the SPIE*, 4836, pp. 154–164.
Kenyon SJ and Bromley BC (2001) Gravitational stirring in planetary debris disks. *Astronomical Journal* 121: 538–551.
Kenyon SJ and Luu JX (1998) Accretion in the early Kuiper Belt: I: Coagulation and velocity evolution. *Astronomical Journal* 115: 2136–2160.
Kenyon SJ and Luu JX (1999) Accretion in the early Kuiper Belt: II: Fragmentation. *Astronomical Journal* 118: 1101–1119.
Kowal CT, Dressler A, Adams R, et al. (1977) 1977 UB (slow-moving object kowal). *IAU Circular 3134*.
Kuiper GP (1951) On the origin of the solar system. In: Hynek JA (ed.) *Proceedings of a Topical Symposium, Commemorating the 50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics*, p. 357. New York: McGraw-Hill.
Lellouch E, Laureijs R, Schmitt B, et al. (2000) Pluto's non-isothermal surface. *Icarus* 147: 220–250.
Levison HF and Stern SA (2001) On the size dependence of the inclination distribution of the main Kuiper Belt. *Astronomical Journal* 121: 1730–1735.
Licandro J, Oliva E, and Di Martino M (2001) NICS-TNG infrared spectroscopy of trans-Neptunian objects 2000 EB173 and 2000 WR106. *Astronomy and Astrophysics* 373: L29–L32.
Licandro J, Pinilla-Alonso N, Pedani M, Oliva E, Tozzi GP, and Grundy WM (2006) The methane ice rich surface of large TNO 2005 FY9: A Pluto-twin in the trans-Neptunian belt? *Astronomy and Astrophysics* 445: L35–L38.
Luu J and Jewitt D (1996) Color diversity among the Centaurs and Kuiper Belt objects. *Astronomical Journal* 112: 2310.
Luu JX, Jewitt DC, and Trujillo C (2000) Water ice in 2060 Chiron and its implications for Centaurs and Kuiper Belt objects. *Astrophysical Journal* 531: L151–L154.

Lykawka PS and Mukai T (2005) Higher albedos and size distribution of large trans-Neptunian objects. *Planetary and Space Science* 53: 1319–1330.

Malhotra R (1993) The origin of Pluto's peculiar orbit. *Nature* 365: 819–821.

Marcialis RL, Rieke GH, and Lebofsky LA (1987) The surface composition of Charon – Tentative identification of water ice. *Science* 237: 1349–1351.

Margot JL, Brown ME, Trujillo CA, and Sari R (2004) HST observations of Kuiper Belt binaries. *AAS/Division for Planetary Sciences Meeting Abstracts* 36.

McBride N, Green SF, Davies JK, et al. (2003) Visible and infrared photometry of Kuiper Belt objects: Searching for evidence of trends. *Icarus* 161: 501–510.

McKinnon WB (2002) On the initial thermal evolution of Kuiper Belt objects. In: Warmbein B (ed.) *Proceedings of Asteroids, Comets, Meteors – ACM 2002, ESA SP-500. International Conference, 29 July–2 August 2002*. Berlin, Germany, pp. 29–38, ISBN 92–9092–810–7. Noordwijk, The Netherlands: ESA Publications Division.

McKinnon WB and Barr AC (2006) Structure and evolution of Ice Dwarf Planets. *AAS/DPS Meeting*, 21.01.

Meech KJ and Belton MJS (1989) (2060) Chiron. *International Astronomical Union Circular* 4770: 1.

Morbidelli A (2006) Origin and dynamical evolution of comets and their reservoirs. *Astro-ph preprint*.

Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.

Noll KS, Stephens DC, Grundy WM, et al. (2002) Detection of two binary trans-Neptunian objects 1997 CQ29 and 2000 CF105, with the Hubble Space Telescope. *Astronomical Journal* 124: 3424–3429.

Orosei R, Coradini A, de Sanctis MC, and Federico C (2001) Collision-induced thermal evolution of a comet nucleus in the Edgeworth–Kuiper Belt. *Advances in Space Research* 28: 1563–1569.

Owen TC, Roush TL, Cruikshank DP, et al. (1993) Surface ices and the atmospheric composition of Pluto. *Science* 261: 745–748.

Pasachoff JM, Souza SP, Babcock BA, et al. (2005) The structure of Pluto's atmosphere from the 2002 August 21 Stellar occultation. *Astronomical Journal* 129: 1718–1723.

Peixinho N, Doressoundiram A, Delsanti A, Boehnhardt H, Barucci MA, and Belskaya I (2003) Reopening the TNOs color controversy: Centaurs bimodality and TNOs unimodality. *Astronomy and Astrophysics* 410: L29–L32.

Prialnik D, Brosch N, and Ianovici D (1995) Modelling the activity of 2060 Chiron. *Monthly Notices of the Royal Astronomical Society* 276: 1148–1154.

Rabinowitz DL, Barkume K, Brown ME, et al. (2006) Photometric observations constraining the size, shape, and albedo of 2003 El61, a rapidly rotating, Pluto-sized object in the Kuiper Belt. *Astrophysical Journal* 639: 1238–1251.

Roe HG (2006) Methane in Pluto's atmosphere. *AAS/DPS Meeting Abstracts* 31.07.

Sicardy B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.

Sicardy B, Widemann T, Lellouch E, et al. (2003) Large changes in Pluto's atmosphere as revealed by recent stellar occultations. *Nature* 424: 168–170.

Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.

Spencer JR, Stansberry JA, Grundy WM, and Noll KS (2006) A Spitzer IRS survey of hydrated and non-hydrated M asteroids: Preliminary 5–13 micron results. *AAS/DPS Meeting*, 34.01.

Spencer JR, Stansberry JA, Trafton LM, Young EF, Binzel RP, and Croft SK (1997) Volatile transport, seasonal cycles, and atmospheric dynamics on Pluto. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, p. 435. Tucson, AZ: University of Arizona Press.

Stansberry JA, Cruikshank DP, Grundy WM, et al. (2004) Far-IR photometry of Centaurs and Kuiper Belt objects with Spitzer. *AAS/Division for Planetary Sciences Meeting Abstracts*, *36*.

Stansberry JA, Grundy WM, Margot JL, et al. (2006) The albedo, size, and density of binary Kuiper Belt object (47171) 1999 TC36. *Astrophysical Journal* 643: 556–566.

Stansberry JA, Lunine JI, Hubbard WB, Yelle RV, and Hunten DM (1994) Mirages and the nature of Pluto's atmosphere. *Icarus* 111: 503–513.

Stephens DC and Noll KS (2005) The high fraction of binaries in the cold classical Kuiper Belt. *DPS Meeting #*37, 56. 16. *American Astronomical Society*.

Stern SA (1989a) ISM-induced erosion and dynamical drag in the Oort cloud. *Bulletin of the American Astronomical Society* 21: 941.

Stern SA (1989b) Pluto—Comments on crustal composition, evidence for global differentiation. *Icarus* 81: 14–23.

Stern SA (1991) On the number of planets in the outer solar system – Evidence of a substantial population of 1000-km bodies. *Icarus* 90: 271–281.

Stern SA (1992) The Pluto–Charon system. *Annual Review of Astronomy and Astrophysics* 30: 185–233.

Stern SA (1995) Collisional time scales in the Kuiper disk and their implications. *Astronomical Journal* 110: 856–868.

Stern SA (1996) On the collisional environment, accretion time scales, and architecture of the massive, primordial Kuiper Belt. *Astronomical Journal* 112: 1203–1211.

Stern SA (2002) Evidence for a collisional mechanism affecting Kuiper Belt object colors. *Astronomical Journal* 124: 2297–2299.

Stern SA (2003) The evolution of comets in the Oort cloud and Kuiper Belt. *Nature* 424: 639–642.

Stern SA (2005) Regarding the accretion of 2003 VB_{12} (Sedna) and like bodies in distant heliocentric orbits. *The Astronomical Journal* 129: 526–530.

Stern SA (in press) New horizons: NASA's Pluto–Kuiper Belt missions. *Space Science Reviews*.

Stern SA, Buie MW, and Trafton LM (1997) HST high-resolution images and maps of Pluto. *Astronomical Journal* 113: 827.

Stern SA and Cheng A (2002) NASA plans Pluto–Kuiper Belt mission. *EOS* 83: 108–109.

Stern SA and Colwell JE (1997a) Accretion in the Edgeworth–Kuiper Belt: Forming 100–1000 km radius bodies at 30 AU and beyond. *Astronomical Journal* 114: 841–849.

Stern SA and Colwell JE (1997b) Collisional erosion in the primordial Edgeworth–Kuiper Belt and the generation of the 30–50 AU Kuiper gap. *Astrophysical Journal* 490: 879–882.

Stern SA and Trafton L (1984) Constraints on bulk composition, seasonal variation, and global dynamics of Pluto's atmosphere. *Icarus* 57: 231–240.

Stern SA and Trafton L (in press) On the atmosphere of objects in the Kuiper Belt. In: Barrucci A, Morbidelli A, and Cruikshank D (eds.) *Trans-Neptunian Objects*. LPI Press.

Stern SA, Trafton LM, and Gladstone GR (1988) Why is Pluto bright? Implications of the albedo and lightcurve behavior of Pluto. *Icarus* 75: 485–498.

Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.

Stern SA, Weintraub DA, and Festou MC (1993) Evidence for a low surface temperature on Pluto from millimeter-wave thermal emission measurements. *Science* 261: 1713.

Strazzulla G and Johnson RE (1991) Comets in the post-Halley era. In: Newburn RL Jr. (ed.). *Irradiation Effects on Comets and Cometary Debris*, vol. 1, pp. 243–275. The Netherlands: Kluwer Academic Publisher.

Strobel DF, Zhu X, Summers ME, and Stevens MH (1996) On the vertical thermal structure of Pluto's atmosphere. *Icarus* 120: 266–289.

Sykes MV, Cutri RM, Lebofsky LA, and Binzel RP (1987) IRAS serendipitous survey observations of Pluto and Charon. *Science* 237: 1336–1340.

Tegler SC and Romanishin W (1998) Two distinct populations of Kuiper Belt objects. *Nature* 392: 49–51.

Tegler SC and Romanishin W (2000) Extremely red Kuiper-Belt objects in near-circular orbits beyond 40 AU. *Nature* 407: 979–981.

Tegler SC and Romanishin W (2003) Resolution of the Kuiper Belt object color controversy: Two distinct color populations. *Icarus* 161: 181–191.

Thébault P and Doressoundiram A (2003) Colors and collision rates within the Kuiper Belt: Problems with the collisional resurfacing scenario. *Icarus* 162: 27–37.

Tholen DJ and Buie MW (1997) Bulk properties of Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*. Tucson, AZ: University of Arizona Press, 193 pp.

Tholen DJ, Buie MW, Storrs AD, and Lark N (1986) Improved physical parameters for the Pluto–Charon system. *Bulletin of the American Astronomical Society* 18: 821.

Thomas N, Eggers S, Ip W-H, et al. (2000) Observations of the trans-Neptunian objects 1993 SC and 1996 TL66 with the infrared space observatory. *Astrophysical Journal* 534: 446–455.

Thompson RW, Murray BG, Khare BN, and Sagan C (1987) Coloration and darkening of methane clathrate and other ices by charged particle irradiation: Applications to the outer solar system. *Journal of Geophysical Research* 92: 14933–14947.

Throop HB (2000) *Light Scattering and Evolution of Protoplanetary Disks and Planetary Rings*. PhD Thesis, University of Colorado at Boulder.

Trafton L (1990) A two-component volatile atmosphere for Pluto. I: The bulk hydrodynamic escape regime. *Astrophysical Journal* 329: 512–523.

Trafton L and Stern SA (1983) On the global distribution of Pluto's atmosphere. *Astrophysical Journal* 267: 872–881.

Trafton LM, Hunten DM, Zahnle KJ, and McNutt RL Jr., (1997) Escape processes at Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*. Tucson, AZ: University of Arizona Press, 475 pp.

Trilling DE and Bernstein GM (2006) Light curves of 20–100 km Kuiper Belt objects using the Hubble Space Telescope. *Astronomical Journal* 131: 1149–1162.

Trujillo CA and Brown ME (2002) A correlation between inclination and color in the classical Kuiper Belt. *Astrophysical Journal* 566: L125–L128.

Trujillo CA, Brown ME, Barkume KM, Schaller EL, and Rabinowitz DL (2006) The surface of 2003EL61 in the near infrared. *Astrophysical Journal* 655: 1172–1178.
Tryka KA, Brown RH, Chruikshank DP, Owen TC, Geballe TR, and Debergh C (1994) Temperature of nitrogen ice on Pluto and its implications for flux measurements. *Icarus* 112: 513–527.
Tyson JA (2002) Large synoptic survey telescope: Overview. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries. Proceedings of the SPIE*, 4836, pp. 10–20.
Veillet C, Parker JW, Griffin I, et al. (2002) The binary Kuiper-Belt object 1998 WW31. *Nature* 416: 711–713.
Walker AR (1980) An occultation by Charon. *Monthly Notices of the Royal Astronomical Society* 192: 47P–50P.
Walker MF and Hardie R (1955) A photometric determination of the rotational period of Pluto. *Publications of the Astronomical Society of the Pacific* 67: 224.
Ward WR (1996) Planetary accretion. Completing the inventory of the solar system. *Astronomical Society of the Pacific Conference Proceedings* 107: 337–361.
Weaver HA, Stern SA, Mutchler MJ, et al. (2006) Discovery of two new satellites of Pluto. *Nature* 439: 943–945.
Womack M and Stern SA (1999) Detection of carbon monoxide in (2060) Chiron. *Solar System Research (Astronomicheskii Vestnik)* 33: 187.
Yelle RV and Elliot JL (1997) Atmospheric structure and composition. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 347–390. Tucson, AZ: University of Arizona Press.
Yelle RV and Lunine JI (1989) Evidence for a molecule heavier than methane in the atmosphere of Pluto. *Nature* 339: 288–290.
Young EF, Binzel RP, and Crane K (2001) A two-color map of Pluto's sub-Charon hemisphere. *Astronomical Journal* 121: 552–561.
Young EF, Buie MW, French RG, et al. (2006) The detailed vertical structure of Pluto's atmosphere from the 12 Jun 2006 Stellar occultation. *AAS/DPS Meeting*, 31.04.
Young LA, Elliot JL, Tokunaga A, de Bergh C, and Owen T (1997) Detection of gaseous methane on Pluto. *Icarus* 127: 258–270.

Relevant Websites

http://www.boulder.swri.edu – The *Distant EKOs* Newsletter, Department of Space Studies of the Southwest Research Institute.
http://www.lsst.org – Large Synoptic Survey Telescope.
http://pan-starrs.ifa.hawaii.edu/public/ – Pan-STARRS: Panoramic Survey Telescope & Rapid Response System.
http://pluto.jhuapl.edu – *New Horizons* Website.
http://taos.asiaa.sinica.edu.tw – Census of Kuiper Comets: The Taiwan–America Occultation Survey.

This page intentionally left blank

10.21 Exoplanetary Geophysics: An Emerging Discipline

G Laughlin, University of California at Santa Cruz, Santa Cruz, CA, USA
JJ Lissauer, NASA Ames Research Center, Moffett Field, CA, USA

Published by .

10.21.1	**Introduction**	673
10.21.2	**Planetary Detection Methods**	674
10.21.3	**An Overview of the Current Planetary Catalog and Census**	677
10.21.4	**Planetary Geophysics**	680
10.21.4.1	Exoplanetary Systems	680
10.21.4.2	The Radii, Densities, and Compositional Structures of Extrasolar Planets	681
10.21.4.3	Observations of Exoplanetary Atmospheres and their Interpretation	684
10.21.5	**Planet Formation**	687
10.21.5.1	An Overview Informed by the Solar System	687
10.21.5.2	How Did Short-Period Planets Arrive at their Observed Locations?	690
10.21.6	**Questions for the Next 10 Years and Beyond**	691
References		692

10.21.1 Introduction

Until nearly the close of the twentieth century, the search for planets beyond the solar system was a fraught and generally less-than-fully respectable endeavor. For generations, the scientific community was subjected to spurious announcements of discovery of planets orbiting nearby stars. Famous examples include erroneous claims, starting in 1855, of a planetary mass object in the 70 Ophiuchi system (e.g., Jacob, 1855; See, 1896) and of one, or possibly two, Jupiter-like planet(s) harbored by Barnard's Star, the red dwarf that is our Sun's nearest neighbor aside from the Alpha Centauri triple-star system (van de Kamp, 1963, 1969).

If we adopt the frequently used definition that requires the planetary mass, M_p, to lie below the $13M_J$ deuterium-burning limit, where M_J denotes the mass of Jupiter, then it might be appropriate to credit the first bona fide detection of an extrasolar planet to Latham et al. (1989), who used Doppler radial velocity (RV) measurements to sense an object with a minimum mass of 11 Jupiter masses on an eccentric 83.9-day orbit around HD 114762, a relatively nearby, solar-type star. With hindsight, this object belongs to the class of giant planets that tend to be more massive than Jupiter, which often have substantial orbital eccentricities and which are found in orbit around ~5% of local solar-type dwarfs (Butler et al., 2006). At the time of HD 114762 b's discovery, however, as a consequence of the abysmal track record of extrasolar planet claims, Latham et al.'s announcement did not generate significant excitement within the wider community. Indeed, the profound difference between HD 114762 b's mass and orbit from those of our solar system's planets led Latham et al. to state, "This leads to the suggestion that the companion is probably a brown dwarf, and may even be a giant planet." Given that HD 114762 b's product of its mass and sine of the inclination angle between its orbital planes, $M_p \sin i$, is so near this physical mass boundary and $\sin i < 1$, Latham et al.'s statement can still be considered as a proper assessment of our current knowledge of this object a quarter century later!

Soon thereafter, astronomers developed a sudden and abiding interest in extrasolar planets. The catalyst was the pulse-timing detection, by Wolszczan and Frail (1992), of two terrestrial-mass planets, with $M_p \sin i = 3.4M_\oplus$ and $2.8M_\oplus$ and with Mercury-like periods of 66.5 and 98.2 days, orbiting the millisecond pulsar PSR 1257 + 12. Pulsars, which are rapidly rotating, strongly magnetized neutron stars, emit radio waves that appear as periodic pulses to an observer on Earth. The pulse period, Doppler shifted to the frame of motion of the barycenter of our solar system, can be determined very accurately. Although a pulsar emits an exquisitely periodic signal, the times at which the pulses reach the receiver are not equally spaced if the distance between the pulsar and the telescope varies in a nonlinear fashion, and if periodic variations are present in data that have been properly reduced, they may indicate the presence of companions orbiting the pulsar. In the case of PSR 1257 + 12, measurements are so accurate that deviations from Keplerian orbits of the two planets mentioned in the preceding text caused by their mutual interactions have been measured and used to calculate their actual masses, $M_p = 4.3 \pm 0.2 M_\oplus$ and $3.9 \pm 0.2 M_\oplus$, implying that the system is inclined by ~50° from the line of sight (Konacki and Wolszczan, 2003). A handful of other pulsar planets have been found, including a lunar-mass third planet orbiting PSR 1257 + 12 and a long-period Jupiter-mass planet in a circumbinary orbit about a pulsar-white dwarf binary system (Sigurdsson et al., 2003).

The pulsar planets were followed in 1995 by Michel Mayor and Didier Queloz's discovery of 51 Pegasi b, a gas giant planet with a minimum mass roughly half that of Jupiter and a startlingly short 4.23077-day orbital period about a metal-rich but otherwise ordinary G-type main-sequence star. ('Official' exoplanetary names are a compound of the name of the parent star and a lower-case letter. The first planet detected in orbit around a given star is denoted *b*, the second, *c*, etc.). Thereafter, the pace of discovery ballooned, along with our understanding both of the planetary census and of the physical characteristics of alien worlds. The study of exoplanet

characteristics has largely been made possible by the detection of transiting planets, which partially eclipse the light of their star once per orbit. The first transiting planet was discovered in 1999 in the orbit around HD 209458 (Charbonneau et al., 2000; Henry et al., 2000).

It is no exaggeration to say that a fully independent branch of astronomy has been created from scratch. The rapid pace of progress is reflected in Figure 1, which charts $\log_{10} M_p \sin i$ as a function of the year of discovery. The lower envelope of the discovery masses in this semilog plot is well described by a constant slope, indicative of a Moore's law behavior with a $\tau \sim 2$-year halving time of the lowest-known extrasolar planetary mass.

10.21.2 Planetary Detection Methods

Given the inherent difficulties associated with the detection of extrasolar planets, it is genuinely remarkable that a very broad range of observational methods for characterization and detection either have been successfully employed or are being seriously prepared for future use. As a corollary to the vast interstellar distances, all extrasolar planets are extremely faint; to a distant optical observer, Jupiter's brightness, $L_J \approx 10^{-7}$, is the solar value.

Most detection methods, therefore, are *indirect*, in that the planets are sensed through their influence on the host star or on the path of light emitted by a distant star. A range of distinct observation methods are sensitive to different classes of planets, and importantly, they provide us with complementary information about the planets they do find. Without question, a full quiver of techniques is required in order to provide an optimal characterization of the diversity of planets and of planetary systems.

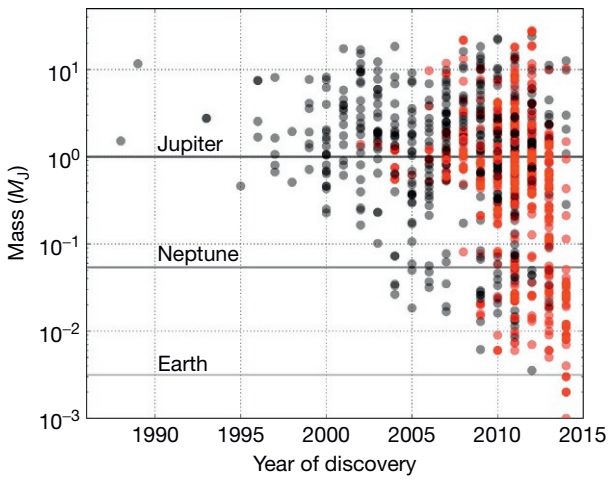


Figure 1 Discovery of extrasolar planets with discovery date versus estimated mass for extrasolar planets orbiting normal stars (not pulsars) over time, with a logarithmic *y*-axis showing mass (or for the cases of Doppler detections, $M \sin i$), and time (in the form of the discovery date) running linearly along the *x*-axis. The lower envelope of this semi-log plot shows a clear Moore's law-like progression and demonstrates that the field has collectively arrived (in the past 1–2 years) at the discovery of planets with $\sim 1 M_\oplus$. Transiting planets are shown in red, whereas planets detected only via Doppler radial velocity are shown in gray. Data are taken from www.exoplanets.org database table, accessed 2/2/2014.

To date, two methods – photometric transit detection and Doppler RV measurement – have generated the vast majority of the successful detections. The Doppler RV technique hinges on the ability to monitor tiny time-dependent wavelength displacements among the profusion of atomic lines in a star's spectrum. By measuring the net overall shift, $z = \delta\lambda/\lambda$, using a large number of individual spectral lines, the velocity at which the star is moving toward or away from an observer can be precisely determined. Astronomers then subtract the motion of the telescope relative to the barycenter of the solar system and other known motions, revealing the radial motion of the target star that results from the tug of its own planets. Transit photometry identifies planet candidates by the periodic drops in observed stellar brightness that are caused by the planet obscuring a portion of the stellar disk once per orbit. Note that transit photometry can only detect planets whose orbits are viewed nearly edge on.

Exoplanet detection using either the RV method or the transit photometry method requires very precise measurements of light from the planet-hosting star. The techniques used to achieve such high precision are based on differential measurements of observations taken at different times. Absolute accuracies, therefore, need not be particularly high, so the mean velocity (relative to the solar system) of a star hosting RV-detected planets is generally not known to an accuracy approaching 1 ms^{-1}, nor is the spectrophotometric brightness of a star hosting a planet observed via transit photometry known to the, for example, part per thousand level needed in relative brightness to detect Neptune-sized planets.

As with pulsar timing, RV measurements yield the product of the planet's mass (divided by that of the star, whose mass can usually be estimated fairly accurately from its spectral characteristics) and the sine of the inclination angle between the orbital plane and the plane of the sky, $M_p \sin i$. The total half-amplitude, *K*, of the stellar reflex velocity induced by an orbiting planet is the measured quantity and is related to $M_p \sin i$ through

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_p \sin i}{(M_\star + M_p)^{2/3}} \frac{1}{\sqrt{1-e^2}} \qquad [1]$$

indicating that the RV technique is most sensitive to massive planets and to planets in short-period orbits. The RV technique allows estimation of the period, *P*; orbital eccentricity, *e*; the mean anomaly, $M = 2\pi(T - T_{\mathrm{periastron}})/P$; and orientation, generally expressed as the longitude of periapse, ω, of the stellar orbit relative to the system center of mass. Multiparameter estimation techniques have reached a high degree of refinement for the problem of determining best-fit orbital elements and uncertainty estimates. For reviews that cover the process of RV modeling, see Ford (2006), Meschiari et al. (2009), Dawson and Fabrycky (2010), and Dumusque et al. (2011). (We recommend the *Systemic Console* software package, available at www.oklo.org, as an intuitive framework for exploring RV and transit-timing data.)

Leading research groups are now routinely achieving instrumentally based velocity errors of $a\ \sigma \sim 50\,\mathrm{cm\,s}^{-1}$ (representing a Doppler shift of two parts in 10^9) on stable, bright,

chromospherically inactive stars. For context, Jupiter causes the Sun's velocity to vary with a half-amplitude $K_J = 12.5\ \mathrm{ms}^{-1}$ and a period $P_J = 11.86$ years; Saturn's effect is the next largest, with an amplitude of $2.7\ \mathrm{ms}^{-1}$ and a period, P_S, of almost 30 years, but the Earth induces a mere $K_\oplus = 6\,\mathrm{cm\,s}^{-1}$ fluctuation. To date, the planet with the lowest RV amplitude signal yet announced is Alpha Centauri Bb, with $K = 51\,\mathrm{cm\,s}^{-1}$, $P = 3.2357$ days, and $M \sin i = 1.1 M_\oplus$ (Dumusque et al., 2012). Earth-like planets orbiting Sun-like stars at astrobiologically interesting distances (where liquid water could be stable on their surfaces) remain somewhat beyond the capabilities currently achieved.

Precise RV measurements require the good statistics that come from aggregating individual measurements of a large number of spectral lines. High-precision measurements thus cannot be achieved for hotter stars with temperatures $T\mathrm{eff} > \sim 6200$ K (corresponding to spectral types O, B, A, and early F), as these stars have far fewer spectral features than do stars of solar type or cooler. Stellar rotation and intrinsic variability (including star spots) also represent major sources of noise that confound RV measurements. The RV surveys, such as the Geneva Extrasolar Planet Search (Mayor et al., 2009), the California Planet Search (Howard et al., 2011), and the Lick–Carnegie Exoplanet Survey (Vogt et al., 2010), are therefore increasingly focusing on very nearby, old, photometrically quiet late G-, K-, and M-type main-sequence dwarfs. The 451 stars that constitute the HARPS Guaranteed Time Observations high-precision sample (Sousa et al., 2008) encompass most of the Sun-like stars at declination $\delta < 20°$ N that are currently receiving intense surveillance.

Planets can also be detected via the wobble that they induce in their parent star's motion projected onto the plane of the sky. This so-called *astrometric* technique is most sensitive to massive planets orbiting stars that are relatively close to Earth. Because the star's motion is detectable in two dimensions, the planet's actual mass, rather than just the combination $M \sin i$, can potentially be measured. Planets on more distant orbits are ultimately easier to detect using astrometry because the amplitude of the star's motion is larger, but finding these planets requires a longer timeline of observations due to their greater orbital periods. To date, despite decades of optimistic expectations (see, e.g., Black and Scargle, 1982), the astrometric method has yet to unambiguously detect an exoplanet. This situation is likely to change soon, however. Simulations by Casertano et al. (2008) indicate that ESA's *Gaia* mission, launched on 19 December 2013 and designed to achieve 7μ as precision on $V = 10$ stars, should be capable of detecting several thousand giant planets out to 3–4 AU from stars lying within 200 pc of the Sun.

The Doppler and astrometric techniques measure a planet's gravitational pull on its star and are thus sensitive to the planet's mass. By contrast, transit photometry detects the amount of starlight that a planet can obscure and thus yields an estimate of the planet's size. If the line of sight from the parent star to the Earth lies in or near the orbital plane of an extrasolar planet, that planet passes in front of the disk of its star once each orbit as viewed from Earth and periodically blocks a small fraction of the star's light. Sufficiently precise photometric time-series measurements of the star's brightness are used to reveal transits, which can be distinguished from rotationally modulated star spots and intrinsic stellar variability by their periodicity, their effectively square-well shapes, and their relative spectral neutrality. Although geometric considerations limit the fraction of planets detectable by this technique (planetary orbital angular momentum vectors are aligned apparently randomly with respect to the galactic plane; the angular momentum of the solar system ultimately derives from the level of turbulent motions that existed within the Sun's parent giant molecular cloud and is not a direct consequence of the differential rotation of the galactic disk), thousands of stars can be surveyed within the field of view of one telescope, so transit photometry is potentially quite efficient. Indeed, transit photometry now generates a clear majority of new planetary candidates. Ground-based surveys, for example, the HATNet Exoplanet Survey (Bakos et al., 2012) and SuperWASP (Smith et al., 2013) survey, are generating dozens of confirmed planets per year.

The preliminary version of the *Kepler* Q12 planet candidate catalog, consisting of signals identified using the first 3 years of *Kepler* data, contains 3538 candidate planets, at least ~90% of which are expected to represent true exoplanets (Casertano et al., 2008; Lissauer et al., 2014a; Morton and Johnson, 2011; Santerne et al., 2013). Almost half of these objects are associated with target stars having two or more planet candidates, implying that configurations of tightly spaced, highly coplanar planets are ubiquitous (Lissauer et al., 2011b). More than 90% of these candidates have estimated radii $R_p < 6R_\oplus$, and almost half have radii $R_p < 2R_\oplus$ despite observational biases against detecting smaller planets. (For an up-to-date list of *Kepler* candidates, see http://exoplanetarchive.ipac.caltech.edu/cgi-bin/ExoTables.) In many of *Kepler's* multiple-transiting planet systems, the transit events exhibit deviations from strict periodicity. These so-called transit-timing variations (as well as transit duration variations that are, much more rarely, also observed) can be attributed to planet–planet interactions within the systems and have, in many cases, established to high confidence that a particular candidate system is the genuine item rather than a false positive. Scores of *Kepler* candidates, furthermore, have been confirmed with Doppler follow-up measurements, for example, Cochran et al. (2011) and Marcy et al. (2014), and hundreds of candidates have been validated to be planets at very high confidence through statistical arguments (e.g., Rowe et al., 2014; Torres et al., 2011).

Reaching further into the box of astrophysical tricks, one finds the microlensing method, which has been successfully employed to investigate the distribution of faint stellar and substellar mass bodies within our galaxy (Alcock et al., 1997) and which has now produced published detections of 18 extrasolar planets (Kains et al., 2013). The microlensing phenomenon stems from the general relativistic bending of the light from a distant star (the source) by a massive object (the lens) passing between the source and the observer. Such lensing can cause the source to appear to gradually brighten by a factor of a few or, for special configurations leading to high-magnification events, factors exceeding 100, over a timescale of weeks to months. If the lensing star has planetary companions, then these far less massive bodies can produce brief enhancements in the brightness – generally of a few hours in duration – provided that the line of sight passes close to the planet. Under favorable circumstances, planets as small as the Earth can be detected. The properties of individual planets, however, and usually even of the stars that they orbit can only be estimated in a statistical sense because of the many parameters

that influence a microlensing light curve (see, e.g., Gaudi, 2012; Peale, 1997). The microlensing technique is comparatively well suited to the detection of planets with long periods and can readily probe Jovian or even super-Earth mass planets in orbits similar to that of Jupiter itself. Due to the prevalence of M-type main-sequence stars in the stellar mass function, the microlensing technique has found a number of examples of giant planets and Neptune-mass planets beyond the locations of the so-called nebular snow lines for these stars (Bennett et al., 2010). Sumi et al. (2011) extensively analyzed several years of microlensing survey data from the galactic bulge and reported the discovery of a population of 'isolated' planetary mass objects with $M \sim M_J$, which appear to occur with somewhat higher frequency than main-sequence stars. Either these planetary mass objects lie in very distant orbits ($d > 100\,\mathrm{AU}$) from undetected parent stars, or they are gravitationally unbound and thus floating freely through the galaxy. The potential presence of hundreds of billions of drifting planetary mass objects, should these worlds withstand statistical scrutiny, would have important consequences for the theory of planet formation and would serve as a definitive endorsement of the microlensing technique.

All of the foregoing are indirect methods to locate extrasolar planets, but what about imaging planets directly? While observational efforts to directly detect extrasolar planets can be traced back to Christiaan Huygens in the 1600s, only very recently have the extreme planet–star brightness contrasts and angular separations been successfully challenged. The reflected starlight from planets with orbits and sizes like those in our solar system is of order 20 visual magnitudes (a factor of $\sim 10^8$) fainter than the stellar brightness, although the brightness contrast improves by a factor of a thousand (or more) in the planet's thermal infrared, for example, at wavelengths $\lambda \sim 10\,\mu\mathrm{m}$. Diffraction of light by telescope optics and atmospheric variability compound the difficulty of direct detection of extrasolar planets. Technological advances, most critically in the area of adaptive optics on large ground-based telescopes and coronagraphic imaging from space, are enabling direct detection of relatively massive, young, and in most cases self-luminous planets orbiting at substantial distances $a \gtrsim 10\,\mathrm{AU}$ from young, nearby stars with masses $M_\star \gtrsim M_\odot$. These detections leverage planet–star brightness contrasts of order $L_p/L_\star < 10^5$ and angular separations of order $\alpha \sim 1$ arcsec. A high-profile example is illustrated by Beta Pictoris b (Lagrange et al., 2010), with mass $M_p = 8^{+5}_{-2} M_J$ at separation $a = 8^{+1.7}_{-0.4}$ AU from Beta Pictoris, a young A-type star that has long been known to harbor an infrared excess and an extensive debris disk (Smith and Terrile, 1984). Another interesting and somewhat controversial example is provided by Fomalhaut b (Kalas et al., 2008). This curious object, which has been imaged with HST, lies near the inner radius of the dust ring surrounding the young, nearby A-type star Fomalhaut. The imaged light associated with the planetary candidate's position is dust-scattered starlight, associated, perhaps, with a circumplanetary disk (Galicher et al., 2013; Kenyon et al., 2014).

The most startling and remarkable direct-imaging detection to date is that of the system of four companions to the young ($\sim 30\,\mathrm{My}$) A-type star HR 8799 (Marois et al., 2008, 2010). These planets all have masses in the vicinity $M_p \sim 5 - 7 M_J$, with semimajor axes lying at 14.5, 24, 38, and 68 AU, and call to mind a version of our own outer solar system that has been scaled up by a factor of several in both mass and astrocentric distance. The presence of HR 8799's planets at their observed distances presents interesting challenges to both the core accretion (Lissauer et al., 2009) and the gravitational instability (Durisen et al., 2007) models of giant planet formation. The proximity of the system configuration to the dynamical stability boundary is also intriguing and implies either that the system will suffer catastrophe within a time that is much less than the star's main-sequence life or that the planets are dynamically protected by a resonant mechanism (Fabrycky and Murray-Clay, 2010). The pace of detection by direct imaging is proceeding at a rapid clip, with additional examples being those of HD 95086b (Rameau et al., 2013), 2MASS J01033563-5515561(AB)b (Rameau et al., 2013), κ And b (Carson et al., 2013), and GJ504b (Kuzuhara et al., 2013), which tend to be massive planets $M \gtrsim 4 M_J$ orbiting at $R \gtrsim 40\mathrm{AU}$ from young (or relatively young) stars.

Direct imaging is poised for major advances with the advent of the Gemini Planet Imager (GPI), which experienced first light at the Gemini South telescope in late 2013 (Macintosh et al., 2014), and by the SPHERE high-contrast adaptive optics imager at the VLT, which came on line in 2014 (for details on the SPHERE instrument, see Beuzit et al., 2008). The GPI adaptive optics system is designed to detect planetary companions at separations of $\delta = 0.2$–1.0 arcsec and brightness contrast of 10^{-7} in 1–2 h exposures. This observational strategy is sensitive to Jupiter-mass planets within the first several million years of their formation and will enable direct imaging of a variety of planets younger than 1 Gy in age. The GPI system, furthermore, uses an integral field spectrometer to obtain $R \sim 40$ spectra that will give important clues to overall atmospheric compositions and states (see, e.g., Konopacky et al. (2013), who reported on the use of the Keck Observatory's OSIRIS integral field spectrograph to detect CO and H_2O absorption in HR 8799c).

Several other methods can also be used to detect extrasolar planets. Precise timing of eclipses of eclipsing binary stars has the potential of revealing the masses and orbits of unseen companions, and indeed, such timing measurements have been used to confirm the planetary nature of the transiting circumbinary planet Kepler-16b (Doyle et al., 2011). Spectroscopy could be used to identify gases that would be stable in planetary atmospheres but not in stars, and Doppler variations of such signals could yield planetary orbital parameters. As an example, Brogi et al. (2013) had obtained tentative detections of CO and H_2O in the emission spectrum of 51 Peg during superior conjunction of its nontransiting short-period planetary companion. Starlight scattered from a planetary atmosphere will be polarized (for detail, see Seager et al., 2000), and the expected polarization signal lies at the threshold of current detectability (Berdyugina et al., 2008; Wiktorowicz, 2009). Radio emissions similar to those detected from Jupiter could reveal the presence of extrasolar planets, although to date all attempts at detection via this channel have been unsuccessful (see, e.g., Lazio et al., 2010). Finally, the detection of artificial signals from an alien civilization would presumably imply the presence of a habitable terrestrial-sized planet or giant planet satellite.

The *Kepler* spacecraft has broken through the psychologically important threshold of one Earth radius, identifying numerous Earth-sized candidates, and has validated several planets smaller than the Earth, including Kepler-37 b, a planet about the size of the Earth's Moon (Barclay et al., 2013). The Doppler RV technique is approaching the comparable $1M_{\oplus}$ threshold, with the Geneva Extrasolar Planet Search team's detection of an $M_p \sin i \sim 1.1M_{\oplus}$ planet in a short-period orbit about Alpha Centauri B (Dumusque et al., 2012). All of the Earth-sized exoplanets verified thus far, however, have orbital periods of a few weeks or less and are not worlds of astrobiological interest. Nevertheless, it has been unambiguously established that planets of masses somewhat greater than the Earth's mass and larger are *extremely* common in the galaxy, and the emphasis of the field is shifting from the detection of extrasolar planets to their physical characterization. In the same manner that a new branch of astronomy has been created, a new and important subfield of planetary geophysics has also come into being.

10.21.3 An Overview of the Current Planetary Catalog and Census

Including both verified exoplanets and high-quality *Kepler*-detected transiting planet candidates, thousands of planets have now been identified, and the outlines of the galactic planetary census are emerging. Much of the current inventory of planets is shown in Figure 2. In this diagram, which combines planets and planet candidates from several sources, the log of the mass ratio of the planet to the central star, $\log_{10}(M_p/M_\star)$, is plotted on the ordinate, and $\log_{10}P$ is charted on the abscissa. For the case of Doppler detections, we conflate M_p with $M_p \sin i$, and for high-quality candidate transiting planets detected photometrically, we have transformed from planetary radius R_p to M_p using a simple transformation:

$$M_p = \left(\frac{R_p}{R_\oplus}\right)^3 M_\oplus \qquad R_p < R_\oplus \qquad [2]$$

$$M_p = \left(\frac{R_p}{R_\oplus}\right)^{2.06} M_\oplus \qquad R_J > R_p > R_\oplus \qquad [3]$$

$$M_p = M_J \qquad R_J > R_\oplus \qquad [4]$$

In the preceding text, the $M_p/M_\oplus = (R_p/R_\oplus)^{2.06}$ relation (Lissauer et al., 2011a) provides a best-fitting power law to the solar system planets spanned in mass by the Earth and Saturn. That such a simple power law does such a good job in the face of different formation mechanisms for the solar system planets can largely be attributed to the fact that only four planets are involved in the fit and to the substantial similarity between Uranus and Neptune. The $M_p = M_J$ relation is applied to just a handful of candidates that appear in multiple-planet systems and does not affect the overall trends exhibited by the diagram. In the absence of solid information, the relation $M_p = (R_p/R_\oplus)^3 M_\oplus$ simply assumes that a small number of planet candidates with radii smaller than the Earth's have Earth-like densities. As we will discuss in the succeeding text, the results from planets that have had both M_p and R_p independently measured indicate that there is no single planetary mass–radius relation and that the dispersion in radii for

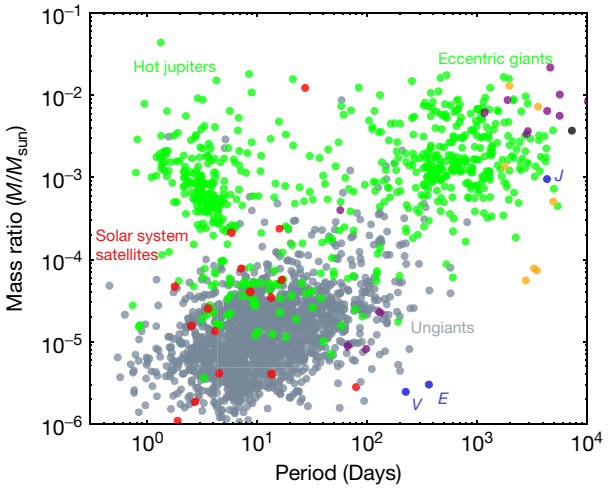


Figure 2 *Green circles*: $\log_{10}(M_{satellite}/M_{primary})$ and $\log_{10}(P)$ for 634 planets securely detected by the radial velocity method, including those first identified via transits. *Red circles*: $\log_{10}(M_{satellite}/M_{primary})$ and $\log_{10}(P)$ for the regular satellites of the Jovian planets in the solar system, along with the Earth's Moon. *Gray circles*: $\log_{10}(M_{satellite}/M_{primary})$ and $\log_{10}(P)$ for 1501 *Kepler* candidates and objects of interest in which *multiple*-transiting candidate planets are associated with a single primary. Radii, as reported in Batalha et al. (2013), are converted to masses using the mass–radius relations described in the text. *Black circle*: Planets detected by direct imaging, with mass ratios estimated from infrared emission detected from the planetary candidate (imaged planets with $P > 10^4$ days, e.g., HR 8799 b, c, d, and e are not shown). *Gold circles*: Planets detected via microlensing. *Purple circles*: Planets detected via pulsar and white dwarf timing. *Blue circles*: Jupiter, Earth, and Venus.

planets of given mass is substantially larger than was generally expected.

The degree of incompleteness that characterizes Figure 2 is somewhat difficult to evaluate, especially in light of the fact that it draws together planets detected by a variety of methods. Nevertheless, a reasonable estimate for the current limiting sensitivity of the Doppler surveys, in the absence of highly focused, high-cadence campaigns on bright nearby stars, is $K = 2\,\mathrm{m\,s^{-1}}$, multiplied by an additional period-dependent term of order unity that favors shorter orbits. A curious and mostly unpredicted feature of the diagram is that the planets fall into three groups. Observational completeness, furthermore, while not perfectly known (see, e.g., Cumming et al., 2008), is sufficient to ensure that the gaps separating the three distinct populations represent real features of the galactic planetary census.

Centered on $M_p/M_\star \sim 1 \times 10^{-3}$ (i.e., $\sim 1M_J$) and $P \sim 3$ days is a population of *hot Jupiters*. The members of this group, on account of their ready detectability, are the best-studied (if still incompletely understood) population of extrasolar planets. In aggregate, hot Jupiters are clearly gas giants dominated by hydrogen–helium compositions, but in nearly every observable property – radii, chemistry, emission properties, density, orbital elements, and so forth – they display a tremendous variability that defies easy categorization.

Hot Jupiters constitute a prominent fraction of the planets in Figure 1 because they are readily detectable, but as a whole, they represent a small fraction of the entire population of

planets. Among a large sample of main-sequence host stars, they are the exception rather than the rule, with an occurrence rate among nearby Sun-like main-sequence stars of only about ~1% (Mayor et al., 2011a; Wright et al., 2012). From the Batalha et al. (2013) catalog of *Kepler* candidate planets, one estimates a hot Jupiter occurrence rate of 0.5%, in contrast to the order-unity occurrence rate of smaller planets with orbital periods $P < 100$ days.

The upper right-hand corner of **Figure 2** shows a second population of planets that have been revealed by the Doppler velocity surveys. In the absence of a consensus terminology, we refer to this group as the *eccentric giants*. Planets in this group are characterized by (1) orbital periods in the range of hundreds to thousands of days; (2) masses that are, on average, several times larger than the mass of Jupiter, with a median value of $M \sin i = 1.94 M_J$, although the mean of the underlying distribution is clearly smaller because more massive planets are easier to detect; and (3) typically substantial eccentricities, with a median value $e = 0.22$ and $e > 0.5$ not uncommon, although a few members travel on nearly circular orbits. HD 114762 b (Latham et al., 1989), with $M \sin i = 11 M_J$ and $P = 114$ days, is a member of this population. Of order $F \sim 5\%$ of the solar types F, G, and K, main-sequence dwarfs in the local galactic neighborhood harbor an eccentric giant, making them roughly ten times more common than hot Jupiters. Like the hot Jupiters, however, the occurrence rate of eccentric giants depends strongly on the metallicity of the parent star (Gonzalez, 1997, 1999). This strong correlation imparts an important, albeit still imperfectly understood, clue regarding the nature of the planet-forming process. The most metal-rich stars, with $[\mathrm{Fe/H}] > 0.3$, that is, more than twice the solar value, are a factor of five to ten times more likely to harbor a hot Jupiter or an eccentric giant than are stars with metallicity less than half that of the Sun (i.e., those with $[\mathrm{Fe/H}] < -0.3$) (Fischer and Valenti, 2005; Santos et al., 2001). Stars with $M_\star > M_\odot$ are also statistically more likely to harbor a hot Jupiter or an eccentric giant.

Jupiter lies near the long-period edge of the currently known population of eccentric giants. Until recently, the natural interpretation of this apparent placement near the tail of the distribution is that it arises from observational incompleteness among genuinely Jovian analogs with Jupiter-like masses and Jupiter-like periods. That is, the apparent dearth of planets with $P \sim 10$ years and $M < 1 M_J$ was thought to be the simple consequence of long required timescales for observation and the difficulty in detecting low-amplitude signals. Such assumptions, however, are increasingly being called into question. As an example, consider the stars in the Keck RV survey as maintained by the Lick–Carnegie Exoplanet Survey (Vogt et al., 2010). As of 2012, the Keck I telescope has been used to obtain 31,910 Doppler measurements in the overall RV database (binned to a minimum cadence of 2 h in the frequently occurring case of multiple back-to-back exposures that average over stellar p-modes) on 1298 individual, primarily solar-type, main-sequence dwarfs. The median-derived instrumental uncertainty on these measurements is $\sigma = 1.45\,\mathrm{ms}^{-1}$ (stellar jitter, depending on the star, generally contributes $\sigma = 1\text{–}5\,\mathrm{ms}^{-1}$ in quadrature to the instrumental uncertainty, and for observations associated with individual stars, the median time baseline $\tau_{\mathrm{base}} = \mathrm{HJD}_{\mathrm{obs.last}} - \mathrm{HJD}_{\mathrm{obs.first}} = 2426$ days). Furthermore, there are 237 stars with time baselines longer than Jupiter's orbital period ($\tau_{\mathrm{base}} > 11.85$ years). The stars in this long-baseline group tend to be similar to the Sun in terms of basic properties and are chromospherically quiet and photometrically stable, averaging 55 Doppler measurements each. A true-Jupiter analog ($P = 11.85$ years and $K = 12 \sin i\ \mathrm{ms}^{-1}$) orbiting almost any of these stars would be immediately evident.

Of potentially equal significance to the concentrations of planets in the mass–period ratio diagram are the regions of parameter space where the number density of worlds is low. Nature tends to avoid sitting giant planets with orbital periods in the 10 days $< P <$ 100 days range; the underlying cause of these broad areas of lower concentration is not fully understood. There appears to be a general dearth of planets of both short and intermediate period with masses ranging from roughly $40 M_\oplus$ to $100 M_\oplus$, consistent with the predictions of rapid growth through this mass range in the core-nucleated accretion model of giant planet formation (Pollack et al., 1996).

Perhaps the most important exoplanet-related discovery of the past 10 years has been the realization that roughly half of the Sun-like stars in the solar neighborhood are accompanied by systems of one or more planets on low-eccentricity orbits with periods ranging from days to months and masses falling in the $1 M_\oplus < M_p < 50 M_\oplus$ range (Mayor et al., 2009, 2011b). These planets are clearly visible as a third population that is distinct from the hot Jupiters and from the eccentric giants. Its members constitute the bulk of the transiting planet candidates in multiple-transiting systems detected by *Kepler*, which are indicated in gray on the mass–period ratio diagram. There is clearly a 'desert' (dearth of objects) between giant planets and these 'ungiants,' the term of art that we will employ for what is largely still *terra incognito* – the planetary mass range $M_p < 30 M_\oplus$.

Whereas the period distribution of giant planets reveals that hot Jupiters are a separate population, distinct from the more distant giants, whose abundance peters out at periods of less than a few months, no such distinction is seen among the ungiants. The population of these planets appears roughly constant in log period space (as would be expected from stability limits imposed by planetary dynamics) for orbital periods between 10 and 100 days. Data are too sparse to extend this conclusion to longer periods. At shorter periods, there is clearly a drop in the abundance of ungiant planets, but this is gradual, with no counterpart to the spike in the 3–4-day range seen for giants. Neptune-sized planets are detected down to orbital periods of 2 days, and planets with $R_p \lesssim 2 R_\oplus$ are observed at periods well under 1 day. The current record holder for ultrashort-period planets is Kepler-78 b, a rocky world (it has been detected both in transit and using RV) slightly larger than the Earth with an orbital period of only 8.52 h (Howard et al., 2013; Pepe et al. 2013).

Little is known about the physical properties of ungiants, apart from that *by mass*, they are composed primarily of elements heavier than helium, in contrast to giant planets, and that they are a diverse lot in terms of dominant constituent *by volume*. It is not clear, for example, whether most of its members are properly described as 'super-Earths' or 'sub-Neptunes' or something else entirely. Curiously, our own solar system, which contains nothing interior to Mercury's 88-day orbit, lacks any of the larger and closer-in planets that dominate this part of the exoplanet diagram; how common terrestrial

planet configurations similar to our solar system are is unknown, although many stars clearly host different and easier-to-detect systems of ungiants Sanchis-Ojeda et al. (2013).

A confounding factor is presented by the fact that very few of the known super-Earth-/sub-Neptune-class planets have been both studied with transit photometry and weighed using Doppler RV and/or transit-timing variations. **Figure 3** shows the current census of planets for which both mass and radius (and hence density) measurements have been made. It provides an interesting high-level synopsis of much of what is currently known regarding the physical properties of extrasolar planets. Even with the log–log scaling, the plot indicates to the eye that planets are subject to a wide variety of structures. For example, within the errors, it appears that planets of $M \sim 6M_\oplus$ range in radii by a factor of at least three. When mass is viewed as a function of radius, the transition in composition appears sharper, with planets with radii $\gtrsim 2R_\oplus$ generally not dominated by rock (including iron) and those with radii $\lesssim 1.6R_\oplus$ and measured masses mostly rocky. While the upper limit to the size of abundant rocky planets within $\sim$0.5AU of their stars seems well established, it is quite possible that many or even most planets smaller than $1.6R_\oplus$ are composed primarily of low-density constituents but that the low masses of such planets make them difficult to measure.

In **Figure 3**, we have overplotted three potential mass–radius relations for ungiants. The first relation is the $M_p/M_\oplus = (R_p/R_\oplus)^{2.06}$ best fit to the solar system planets spanned in mass by the Earth and Saturn. Despite the uncertainties, this relation is seen to provide a reasonable fit to the mass–radius relation of planets smaller in mass than Saturn. The solar system fit is bracketed by two alternate relations. The first, with a steeper slope, is derived from the assumption that low-mass, short-period planets have rocky cores surrounded by hydrogen–helium envelopes, with envelope masses dictated by *in situ* formation, as discussed in Chiang and Laughlin (2013). The second alternate relation, $3M_p/M_\oplus = R_p/R_\oplus$, has a shallower slope and is a fit derived from an ensemble of transit-timing measurements applied to 22 systems within *Kepler*'s multiple-transiting planet candidate population by Wu and Lithwick (2012) (see also Weiss et al., 2013).

Kepler is capable of detecting rocky terrestrial planets analogous to Venus and Earth, but the spacecraft is unlikely to find a statistically robust sample of terrestrial planets similar to those in our solar system. The mass–period diagram is thus still largely incomplete for terrestrial planets with solar system-like masses and orbits and will likely remain so for some time. In the absence of solid data, we can only resort (with an abundance of cautionary disclaimers) to statistical extrapolations to anticipate the inventory of low-mass planets (**Figure 4**).

To construct a concrete (albeit uncertain) extrapolation, we assume that planet formation is highly efficient and that long-distance migration ($\Delta a \gtrsim a$), where a is the current semimajor axis of the planet and Δa is the amount it has migrated, does not regularly occur in the super-Earth population. We then idealize the population of *Kepler* planetary candidates from Batalha et al. (2013) as being composed of solar-composition solids, and we assign each planet a surface density, $\sigma_{solid} = M_p/2\pi a_p^2$,

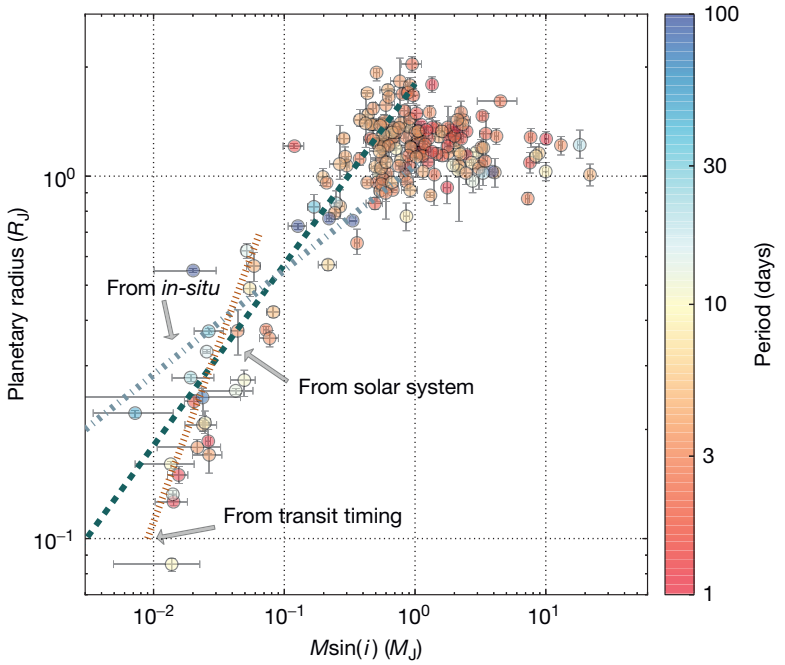


Figure 3 Mass–radius diagram for extrasolar planets. This figure shows all planets for which both mass and radius estimates exist, with orbital period coded to run from 1 day (red) through 100 days (blue). Three sample mass–radius relations are shown. The short dotted line is drawn from a simple *in situ* formation theory as described in Chiang and Laughlin (2013) and is based on rock-iron planets surrounded by hydrogen–helium envelopes and does not include the possible effects of insolation-driven atmospheric mass loss. The long dashed line is the best-fit $M_p/M_\oplus = (R_p/R_\oplus)^{2.06}$ power law appropriate to the solar system planets spanning masses from the Earth through Saturn. The medium dashed line is the $3M_p/M_\oplus = R_p/R_\oplus$ relation derived by Wu and Lithwick (2012) from consideration of the transit-timing variations observed by *Kepler* for 22 multiple-transiting candidate planet pairs in the vicinity of mean motion resonances.

where a_p is the orbital semimajor axis, computed with the approximation that $M_\star = M_\odot$, and $M_p = (R_p/R_\oplus)^{2.06} M_\oplus$. The resulting $N = 1925$ surface densities obtained for low-mass candidates with $R_p < 5R_\oplus$ are plotted in Figure 3. The best-fit power law to the median data can be interpreted to define a 'minimum-mass extrasolar nebula' $\sigma_{solid} = 50(a/1\,AU)^{-1.6}\,g\,cm^{-2}$, which has a very similar power law index, $\alpha = -1.6$ to the canonical minimum-mass solar nebula (MMSN) value of a $\alpha = -1.5$ (Hayashi, 1981), but with a density normalization that is 5 times as large. Extending the power law to a distance $a \sim 1$AU leads to two suggestions or, better, two moderately informed speculations:

- Roughly half of the Sun-like stars in the Sun's neighborhood will harbor a planet as large as (or larger than) the Earth in (or near) an Earth-like orbit.
- Among the planetary systems of single main-sequence stars, the inventory of mass in the solar system's terrestrial planets will lie below the median total planetary mass interior to $a \sim 2$ AU.

10.21.4 Planetary Geophysics

Investigations of extrasolar planets can currently be divided conceptually into (1) analyses of planetary structure and composition, (2) the exploration of planetary orbital dynamics, and (3) the study of planet formation. These three areas closely inform one another. For example, at a given planetary age, τ_p, the radius of a giant planet (a zeroth-order structural property) depends on the mode of formation. A planet that emerges from the process of core accretion will probably be smaller at early τ_p than one that arises from disk gravitational instability (Marley et al., 2007). The planetary radius, in turn, has a strong influence on the tidal dissipation rate, with tidal luminosity obeying $\dot{E} \propto (R_p/a)^5$, and frequently, the long-term dynamical behavior of a compact planetary system is sensitively informed by its tidal evolution, for example, Mardling and Lin (2002). In the remainder of this chapter, we will give an overview of what is known about the 'geophysics' of extrasolar planets, with an eye toward elucidating the connections to theories of dynamical evolution and theories of planetary formation.

The direct characterization of extrasolar planets has been enabled in large part by elaborations on transit photometry, and as a result, our understanding is strongly biased toward short-period planets orbiting close to their parent stars. Most of the well-characterized giant exoplanets were discovered by ground-based transit searches, with most of the ungiants found by the Kepler spacecraft (Lissauer et al. 2014b). With this context, we then provide a synopsis of the theory of planet formation, largely informed by the contents of our own solar system. We argue that a synthesis of the overall picture is most definitely not yet fully in hand, which motivates a look forward at advances that are likely to be made within the next decade.

10.21.4.1 Exoplanetary Systems

RV surveys have detected scores of systems with more than one exoplanet, including more than one dozen nearby stars with three or more planets. In systems with high multiplicity, the planets tend to have low masses and low eccentricities, although a number of systems with three or more giant planets are known. Planets tend to be clustered into systems in the sense that there are more multiple-planet systems than would be the case were planets randomly distributed among stars. Hot Jupiters are an exception to this rule – few of them are known to have companion planets. (Nonetheless, the star υ Andromedae has at least three Jovian-mass planets, one of which is a hot Jupiter.) The remaining planets are far more distant from υ And and travel on eccentric orbits.

The multiple-transiting systems from *Kepler* provide a much larger and very rich dataset of short-period (primarily from a few days to several months) planets that can be used to

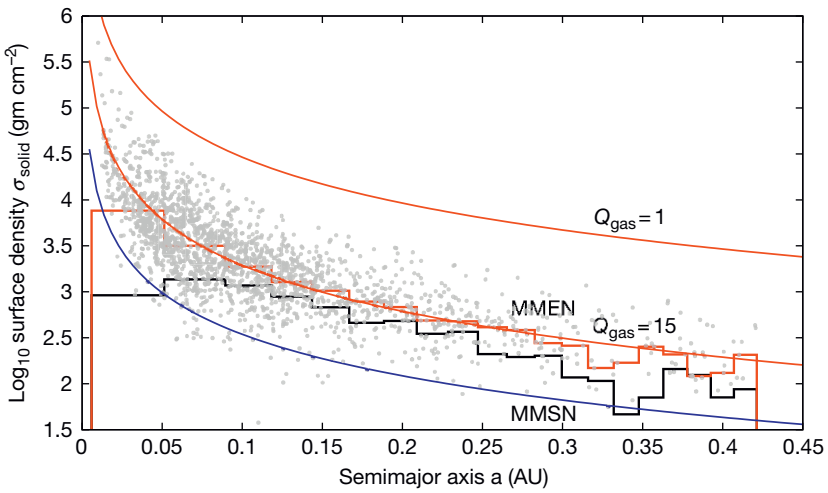


Figure 4 The solid surface density profile of the 'minimum-mass extrasolar nebula' (MMEN) constructed from *Kepler* data (Chiang and Laughlin, 2013). *Gray circles*: $\sigma_{solid,i} \equiv M_i/2\pi a_i^2$, computed from *Kepler* planets with $R_p < 5R_\oplus$ and $P < 100$ days, assuming $M_p/M_\oplus = (R_p/R_\oplus)^{2.06}$ and solar-mass host stars. $Q_{gas} < 1$ implies the nebular surface density in gas required for gravitational instability. For reference, σ_{solid} for the minimum-mass solar nebula (MMSN) – or technically its extrapolation inward, since no planet is present interior to Mercury at $a = 0.4$AU – is plotted as a blue solid curve. As plotted, the MMEN is a factor of five times as massive as the MMSN at these distances.

powerfully test theoretical predictions of the formation and evolution of planetary systems.

The major findings, updated from (Lissauer et al., 2011b), are the following:

1. The large number of candidate multiple-transiting planet systems observed by *Kepler* shows that flat (i.e., nearly coplanar) multiplanet systems are common in short-period orbits around other stars. This result holds for ungiant planets in the size range of $R_p \sim 1-6R_\oplus$, but not for giant planets. Not enough data are yet available to assess its viability for smaller worlds nor for planets with orbital periods longer than a few months.
2. Mean-motion orbital resonances and near resonances are statistically favored over random period ratios, but most planets are neither in nor very near such resonances. First-order resonances (e.g., period ratios of 2:1 and 3:2) dominate, but second-order resonances (with period ratios 3:1, 5:3, etc.) also are manifested. Most near-resonant planet pairs have period ratios from one to a few percent larger than those of the nearby resonance. Few planet pairs have period ratios slightly smaller than those of first-order resonances.
3. Almost all candidate systems survive long-term dynamical integrations that assume circular, planar orbits and the mass–radius relationship given in eqns [2]–[4], derived from planets within our solar system.
4. Attempts to generate a simulated ensemble of planetary systems to match the observed *Kepler* planets do not strongly prefer fits by a single homogeneous population. In particular, there is some evidence for a population or subpopulation of systems that either contains only one detectable planet per star or multiple planets with high relative orbital inclinations that produces the observed excess of singly transiting systems. A single population, whose characteristics are described in the succeeding text, can account for the vast majority of multiple-transiting systems and roughly half of the single planets observed. A third, rarer, group of planar densely packed multiplanet systems also appears to be present. Note that these populations of planetary systems need not be cleanly separated; that is, there may also be significant numbers of intermediate systems.
5. The abundance and distribution of multiply transiting systems imply that the mean number of planets in the $1.5R_\oplus < R_p < 6R_\oplus$ and $3 < P < 125$-day range per star for those stars with at least two such planets of this radius is $\sim$3. According to *Kepler* observations, only $\sim$5% of stars have planetary systems within these size–period limits.
6. The inclination dispersion of most candidate systems with two or more transiting planets appears to have a median of $\sim 2^\circ$ (e.g., Fabrycky et al., 2014), suggesting relatively low mutual inclinations similar to the solar system.
7. Many *Kepler* targets with 1 or 2 identified planet candidates must be multiplanet systems where additional planets are present that are not transiting and/or too small to be detected.

10.21.4.2 The Radii, Densities, and Compositional Structures of Extrasolar Planets

Hot Jupiters, which have $P \lesssim 7$ days and $M_p \gtrsim 100M_\oplus$, represent extrasolar planets at their most alien, but they are nevertheless by far the best studied. The short orbital periods and large masses of these worlds allow for precise and full orbital characterization for cases in which transits occur (transit probabilities are given by $\mathcal{P}_{tr} = 0.0045(1\ \mathrm{AU}/a)(R_\star + R_p/R_\odot)[(1 + e\cos(\pi/2 - \omega))/1 - e^2]$ where ω is the longitude of periastron of the star such that $\omega = 90^\circ$ corresponds to the prospective transit midpoint and e is the orbital eccentricity. In contrast, For a distant observer sited at a random viewing angle, $\mathcal{P}_\oplus = 0.5\%$. $\mathcal{P} \sim 10\%$ for typical hot Jupiters and their high surface temperatures generate infrared planetary fluxes that are readily detectable. Scores of short-period giant planets now have accurate physical measurements, but the spectra and even the radii of these planets have defied straightforward explanation. The origins, furthermore, of the majority of the physically characterizable extrasolar planets are not yet fully clear. It is generally thought that conditions at $a \lesssim 0.1-1$ AU in protostellar disks are not favorable to the successful completion of the core accretion process (Bodenheimer et al., 2000). It is therefore believed that planetary migration, either via quiescent disk processes (Lin et al., 1996; Lubow and Ida, 2011), via planet–planet scattering (Jurić and Tremaine, 2008), or via Lidov–Kozai cycling with tidal friction (Eggleton and Kiseleva-Eggleton, 2001; Fabrycky and Tremaine, 2007), or some combination thereof, plays an important role in delivering the planets to their current locations.

The characterization of extrasolar planets with $M_p < 30M_\oplus$ is difficult, even when the orbital periods are short. Low-mass planets have small transit depths, small RV half-amplitudes, and generally small transit-timing variations, which make it hard to accurately determine their densities. Nevertheless, dedicated follow-up on small-radius transiting planets identified by the *Kepler* mission has begun to yield results (Jontof-Hutter et al., 2014; Marcy et al. 2014), and **Figure 5** shows that the broad outlines of the planetary distribution for $M_\oplus < M_p < 30M_\oplus$ are being delineated. While this diagram is

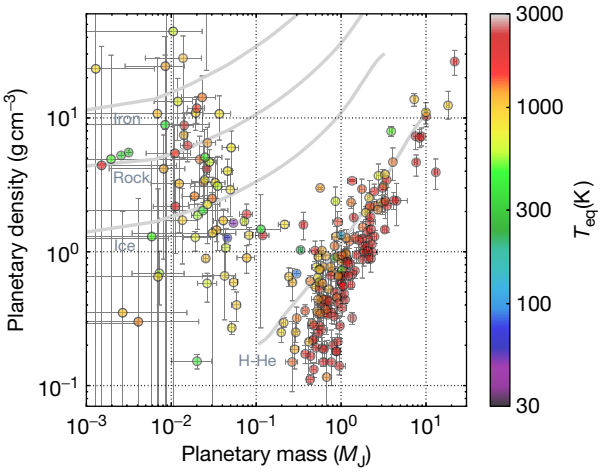


Figure 5 Densities and masses for planets that have individually measured values for R_p and M_p. Theoretical mass–radius relations for planets of various pure compositions are shown for comparison. Note that many of the measurements for low-mass planets were derived from transit-timing variations and these data points tend to have large uncertainties. The six most massive solar system planets (Venus, Earth, Uranus, Neptune, Saturn, and Jupiter) are included on the diagram and can be distinguished on account of their low equilibrium temperatures.

generally consistent with the $\rho(R_p)$ values exhibited by the solar system planets, it is clear that the ungiant regime spanning ($M_\oplus < M_p < 30M_\oplus$) encompasses a very wide range of planetary densities (and by extension, planetary compositions). Frustratingly, there is substantial compositional degeneracy if planetary mass and radius are the only quantities that are measured. Often, a substantial range of admixtures of metallic, rocky, icy, and enveloping gaseous components can jointly contribute to produce a planet of given mass and radius (Adams et al., 2008). The lack of significant concrete physical data for low-mass planets means that guidance must be drawn from theoretical work. The internal structures of planets more massive than Earth are reviewed in Baraffe et al. (2014) and the references therein. The expected properties of 'ocean' planets with large water fractions have been discussed by a number of authors (see, e.g., Sotin et al., 2007, 2011).

Figure 5 suggests that the overall distribution of planetary densities reaches a broad minimum, $\rho_{min} \sim 0.2$ to 1.0 g cm^{-3}, at roughly $M_p \sim 0.1 M_J$. Above this mass, planetary compositions are dominated by hydrogen and helium. Over 130 well-characterized examples of transiting H–He-dominated giant planets (with $M_p/M_\oplus \gtrsim 100$) are now known. It is clear that within this regime, variations in mass and stellar insolation are not the only factors responsible for the observed range in planetary sizes, and *radius anomalies* – observed planetary sizes that are substantially different from those predicted by structural models – have been evident since the discovery of the first transiting extrasolar planets. Early structural evolutionary models for giant planets computed by, for example, Bodenheimer et al. (2001) and Guillot and Showman (2002) suggested that a H–He-dominated planet with HD 209458b's mass, insolation, and age should have a radius of approximately $R_p \approx 1.1 R_J$, a figure that is startlingly at odds with the observed value, $R_p = 1.38 \pm 0.02 R_J$ (Southworth, 2010).

During the last decade, the accumulation of many mass–radius measurements for giant planets has demonstrated that radius anomalies are by no means anomalous (see, e.g., Burrows and Orton (2011) or Baraffe et al. (2010) for overviews). The substantial variation in observed radii for the giant planet population is indicated by Figure 6, which charts the measured radii and uncertainties for the transiting giant planets with accurately measured masses against an insolation-based orbit-averaged planetary equilibrium temperature given by

$$T_{eq} = \left(\frac{R_\star}{2a}\right)^{1/2} \frac{T_\star}{(1-e^2)^{1/8}} \qquad [5]$$

which implicitly assumes that the planet has an albedo of zero and reradiates energy uniformly over its entire surface and depends on the average flux received by the planet over its orbit (incorporating the effect of albedo results in multiplying the right-hand side by $(1-A)^{1/4}$). In eqn [5], $T_\star$ is the stellar effective temperature, e is the planetary orbital eccentricity, $R_\star$ is the host star's radius, and a is the planetary semimajor axis. The actual temperatures in most of the exoplanetary atmospheres are unmeasured and, depending on the atmospheric properties, chemistry, and dynamics, may be higher or lower than T_{eq}.

At a given planetary mass, it is expected that the radius of a mature gas giant planet should primarily be determined by the amount of radiative energy that it receives from its star. This

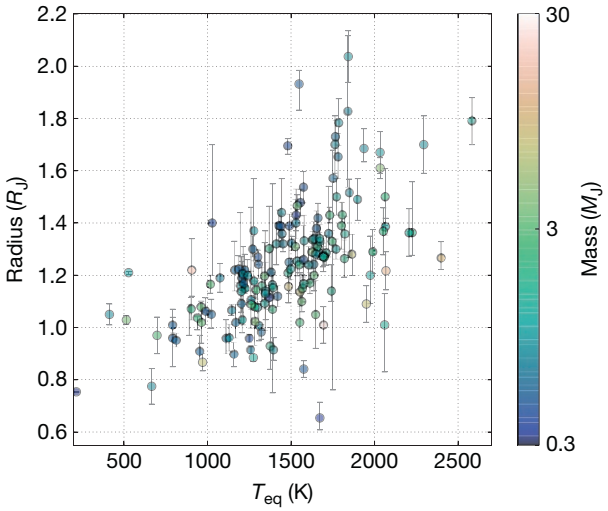


Figure 6 Planetary radii, R_p/R_J, versus insolation-derived planetary equilibrium temperatures, T_{eq} (see eqn [5]), for 130 planets with $M_p \sin i / M_\oplus > 100$ that have both transit and radial velocity characterizations. The planets are color-coded by mass, suggesting that planets substantially more massive than Jupiter are less likely to exhibit significantly inflated radii.

expectation is supported by the clear trend in Figure 6, which indicates that R_p is positively correlated with planetary T_{eq}. The degree of quantitative agreement can be obtained by computing model radii, R_m, for H–He composition planets as a function of M_p and T_{eq} and comparing with those of the observed planets. Figure 7 shows the result of such a comparison to a set of baseline core-free models computed by Bodenheimer et al. (2003) using a Henyey-type planetary structure and evolution code. We define the radius anomaly, $\mathcal{R}$ to be the difference (in Jupiter radii) between the observed planetary radius and the radius predicted by the Bodenheimer et al. (2003) baseline model.

Figure 7 indicates that the structural models explain some, but not all, of the observed variation in planetary radii. For $T_{eq} \gtrsim 1000$ K, planets are usually larger (and sometimes substantially larger) than predicted by the models. At lower temperatures, radii tend to be smaller than what one would expect for a giant planet of solar composition. Several authors, for example, Enoch et al. (2011), have noted the trend that is evident in Figure 7, and a power law fit to the current data suggests an $\mathcal{R} \propto T_{eq}^{1.4 \pm 0.6}$ dependence.

Giant planets with negative $\mathcal{R}$ probably contain a substantial fraction of heavy elements. The majority of observed planets with $T_{eq} < 1000$ K have negative radius anomalies; for this population, $\overline{\mathcal{R}} = -0.12$, which suggests that the average short-period giant planet has a $M_{core} \sim 30 M_\oplus$ core mass. Planets orbiting metal-rich stars are statistically associated with larger core masses (Miller and Fortney, 2011), and there is likely an inverse relation between planetary mass, M_p, and the degree of a planet's enrichment in heavy elements, $Z_p/Z_\star$ (Laughlin et al., 2011).

In order to inflate an H–He-dominated planet to large $\mathcal{R}$, models must appeal to a cryptic source of heating at adiabatic depth, and a wide range of heating mechanisms have been proposed. Those that have received significant attention include dissipation stemming from (1) tidal orbital circularization

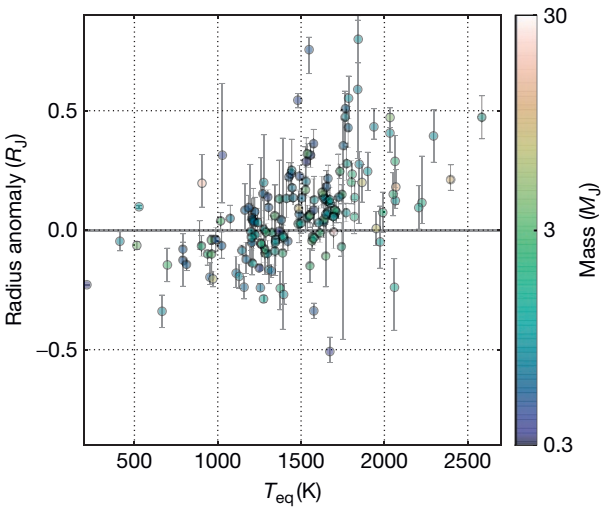


Figure 7 'Radius anomalies,' $\mathcal{R} \equiv R_{obs} - R_{pred}$, versus insolation-based planetary equilibrium temperature, T_{eq}. Planetary models that account for insolation *typically predict radii that are too small* when $T_{eq} > 1200$K.

(Bodenheimer et al., 2001), (2) obliquity tides (Winn et al., 2005) (but see the cautions raised by Levrard et al. (2007) and see Fabrycky et al. (2007) for full-blown arguments against obliquity tides), (3) 'kinetic heating' in which wind energy is converted into heat (Guillot and Showman, 2002), (4) dissipation induced by thermal tides (Arras and Socrates, 2010), and (5) ohmic dissipation (Batygin and Stevenson, 2010).

A clue is provided by the $\mathcal{R} \propto T_{eq}^{1.4}$ dependence. It is difficult to understand this correlation if the radius anomalies stem primarily from tidal heating, for which one would expect the major dependence to be with eccentricity, $\mathcal{R} = \mathcal{R}(e)$. Individual planets, of course, may be significantly affected by dissipation, and there is a wide observed dispersion in planetary size at all T_{eq}.

Arras and Socrates (2010) revived the thermal tidal torque mechanism that was introduced by Gold and Soter (1969) to account for Venus' low-frequency retrograde spin. In Arras and Socrates's (2010) picture, uneven global heating of a short-period planet produces a comparatively cold, high-density bulge that trails the planet–star line. Gravitational torque acting on the bulge spins the planet up into a state of asynchronous rotation, enabling steady-state constant-lag tidal energy dissipation to occur in the planet. Arras and Socrates (2010) suggested that this dissipation may be sufficient to inflate hot Jupiters to their observed radii. Other workers, however, for example, Gu and Ogilvie (2009), argue that rapid isostatic readjustment of the planet will offset the thermal bulges for forcing frequencies (e.g., the planetary mean motion, $n = 2\pi/P$) that are lower than the planetary dynamical frequency (which is of order $\sqrt{\rho/G}$).

The kinetic heating mechanism, as described by Guillot and Showman (2002), posits that a constant fraction, $\eta \sim 0.01$, of the total flux received by a planet is converted into kinetic energy by atmospheric gradients and that this energy is dissipated at adiabatic depth within the planet. While the flow patterns on the surfaces of strongly irradiated extrasolar planets are still a matter of debate (see the succeeding text), an overall $\dot{E} \propto T_{eq}^4$ scaling is implied by the kinetic heating hypothesis, leading to $\mathcal{R} \propto T^{2/3}$, which is somewhat weaker than the observed $\mathcal{R} \propto T^{1.4}$ dependence. As was the case with the evaluation of the viability of tidal heating, this argument does not imply that kinetic heating is absent, but rather, the implication is that kinetic heating is unlikely to be the primary mechanism that is generating the radius anomalies.

Batygin and Stevenson (2010) presented a case that resistive heating (ohmic dissipation) can inflate hot Jupiters. In their model, significant atmospheric electrical conductivity at $T_{eq} \gtrsim 1200$K is enabled by valence electron ionization of alkali metals. The presence of Na in the atmosphere of HD 209458b has been directly observed via transit spectroscopy (Charbonneau et al., 2002). The Batygin–Stevenson model assumes that the atmospheric layers of the swollen planet have rotation-induced zonal jets, which are seen on all four solar system giants and which are a generic feature in two- and three-dimensional hydrodynamic simulations of exoplanetary surface flows (Showman and Polvani, 2011). Conductive atmospheric zonal flows introduce a meridional surface current that, in the Batygin–Stevenson model, flows from the planetary poles to the equator and then closes the circuit by returning poleward through the deep convective interior of the planet. Resistive heating within the planet

$$\dot{E} = \iiint \frac{J^2}{\sigma_r(r)} \mathrm{d}V \qquad [6]$$

where $\sigma_r(r)$ is the conductivity in the near-surface (but still adiabatic) layers, provides an internal energy source. Batygin et al. (2011) found that this heating is sufficient, for reasonable planetary atmospheric and **B**-field values, to provide $\mathcal{R}$ values of order those that are observed.

If one further assumes that the planetary Elsasser number, $\Lambda = \sigma_i B^2 / 2\rho\Omega$, is roughly constant across the aggregate of transiting planets, the Batygin–Stevenson theory implies an $\mathcal{R} \propto T_{eq}^{2.7}$ dependence. This power law index $\alpha = 2.7$ is steeper than the observed value of $\alpha \sim 1.4$. A more sophisticated treatment, however, must account for the back reaction onto the wind velocity, **v**, generated by the Lorentz force:

$$\rho \frac{\mathrm{dv}}{\mathrm{d}t} = \frac{1}{c} \sigma_r (\mathbf{v} \times \mathbf{B}) \times \mathbf{B} \qquad [7]$$

where **B** is the local magnetic field strength and σ_r is the conductivity of the gas parcel. Three-dimensional simulations by Rauscher and Menou (2013) suggest that the Lorentz force provides a drag term to the velocity field, which in turn reduces the current and decreases the dissipation rate, driving the power law index to lower values. In summary, it appears likely that ohmic dissipation may provide the bulk of the explanation for the large radius anomalies that characterize many short-period hot Jupiter-type planets.

The realm of the gas giants has a lower-mass bound of about $50 M_\oplus$. **Figure 5** (which charts masses versus densities) suggests that for $M_p < 50 M_\oplus$, the overall fraction of refractory material increases steadily toward lower masses. This observed trend is in part due to selection effects – the lowest-mass planets for which both R and M can be measured have short periods and therefore very high surface temperatures. As an example, Kepler-78 b (Howard et al., 2013; Pepe et al., 2013) has $M_p \sim 1.7 M_\oplus$, $R_p \sim 1.1 R_\oplus$, and $\rho_p \sim 5.5\,\mathrm{g\,cm^{-3}}$ and must therefore be composed largely of metals and silicates, but its orbital period is only $P = 0.35$ days, and it is so close to its star that it

would have lost any low-density component to its original mass. Indeed, the lower T_{eq}, low $-M_p$ planets shown in Figure 5 whose masses have been determined through analysis of transit-timing variations (and which are thus less sensitive to a bias toward very short periods) have systematically lower densities than those that have had their masses determined via Doppler velocity measurement (Jontof-Hutter et al., 2013).

Bulk densities give an important clue to the overall planetary structures, but, unfortunately, there are a wide range of compositional structures involving admixtures of metals, rocks, ices, and gases that can reproduce a given planetary mass and radius. See, for example, Fortney et al. (2007) for a catalog of planetary structural models spanning a broad range of compositional assumptions.

The tidal Love number, k_2, provides an effective parameterization of the degree of central condensation exhibited by a particular density structure, $\rho(r)$, with lower values for k_2 mapping to higher degrees of central condensation. Specifically, if $\rho_m(r)$ is the mean density of a planet interior to radius r and R_p is the planet's radius, a quantity $\eta_2(R_p)$ can be obtained by integrating the ordinary differential equation

$$r\frac{d\eta_2}{dr}+\eta_2^2-\eta_2-6+\frac{6\rho(r)}{\rho_m(r)}(\eta_2+1)=0 \qquad [8]$$

radially outward from $\eta_2(0)=0$. The Love number k_2 is then calculated as (Sterne, 1939)

$$k_2=\frac{3-\eta_2(R_p)}{2+\eta_2(R_p)} \qquad [9]$$

If a short-period planet in a coplanar, hierarchical ($a_c \gg a_b$) two-planet system has tidally evolved to an *eccentricity fixed point* (see, e.g., Mardling, 2007; Wu and Goldreich, 2002), then the outcome of this dissipative, deterministic process is that the apsidal lines of the two planets are forced to precess at the same rate. To a good approximation, the secular gravitational perturbation from the inner planet, b, fully dictates the precession of the outer planet, c,

$$\dot{\overline{\omega}}_{c\ \mathrm{secular}}=\frac{3}{4}n_c\left(\frac{M_b}{M_*}\right)\left(\frac{a_b}{a_c}\right)^2\frac{1}{\left(1-e_c^2\right)^2}\left[1-\frac{5}{4}\left(\frac{a_b}{a_c}\right)\left(\frac{e_b}{e_c}\right)\frac{1+4e_c^2}{1-e_c^2}\cos\left(\overline{\omega}_b-\overline{\omega}_c\right)\right] \qquad [10]$$

and all of the quantities in the aforementioned expression are potentially measurable through a combination of photometric transit and Doppler-based observations.

The precession rate of the inner planet, however, is dictated by a number of contributions (Batygin et al., 2009):

$$\dot{\overline{\omega}}_{b\,\mathrm{total}}=\dot{\overline{\omega}}_{b\,\mathrm{secular}}+\dot{\overline{\omega}}_{b\,\mathrm{tidal}}+\dot{\overline{\omega}}_{b\,\mathrm{GR}}+\dot{\overline{\omega}}_{b\,\mathrm{rotational}}=\dot{\overline{\omega}}_{c\,\mathrm{secular}} \qquad [11]$$

For an inner planet whose orbital figure, mass, and radius have all been determined via high-quality photometric transit and Doppler-based observations, k_2 constitutes the only unknown in $\dot{\overline{\omega}}_{b\,\mathrm{total}}$. As a consequence, for systems with favorable orbital configurations, one potentially has access to a mechanism that directly probes the interior structure and that can potentially discriminate between 'super-Earth' and 'sub-Neptune' configurations. To date, this method has only been applied to the HAT-P-13 b–c system (Batygin et al., 2009), where the inner planet mass is $M_b=0.85\pm0.038M_J$, well above the $M_p<50M_\oplus$ scale of interest here. Kramm et al. (2012) found $0.265<k_2<0.379$ for HAT-P-13b, which requires that any central core in the planet has mass $M_{core}<27M_\oplus$. The TESS mission (Ricker et al., 2010), which is scheduled for launch in 2017, and the PLATO mission (Barban et al., 2013), selected in February 2014 and planned for launch in 2024, may detect nearby transiting multiple-planet systems for which the orbital properties can be determined with enough precision to enable accurate measurement of k_2. Finally, implicit evidence for the widespread existence of dense planetary cores is provided by large rocky planets such as CoRoT-7b (Léger et al., 2009) and Kepler-10b (Batalha et al., 2011). For a detailed overview of the types of systems for which interior structures can be probed, see Becker and Batygin (2013).

10.21.4.3 Observations of Exoplanetary Atmospheres and their Interpretation

The foregoing discussion makes it clear that the radii and interior structures of the hot Jupiters continue to provide significant challenges to a comprehensive theoretical understanding. An analogous situation also holds for the atmospheric characterization of these planets.

Nevertheless, given the extraordinary difficulties inherent in separating the signal of a faint planet from a bright star, there have been a number of real successes, the majority of which have come from space-based measurements. The Spitzer Space Telescope (Werner et al., 2004) was designed prior to the detection of transiting extrasolar planets, and exoplanetary observations were not part of the mission's scientific requirements. Serendipitously, however, Spitzer has played a key role in probing exoplanetary atmospheres. HST, MOST, CoRoT, and *Kepler* have all also made substantial contributions, as have ground-based telescopes including Keck and the VLT. The literature on the characterization of the atmospheres of short-period planets is now extensive, and space considerations preclude an exhaustive summary here. See Seager and Deming (2010), Perryman (2011), Line et al. (2012), Tinetti et al. (2013), and Madhusudhan et al. (2014) for successively recent overviews. It is important to stress that while the accumulated catalog of measurements is by now impressive, the signal-to-noise ratio of individual measurements is often low; systematic errors may, in general, be significant; and the number of unknowns, including composition, meteorology, and chemistry that characterize an atmosphere, is daunting.

Atmospheric characterization has mostly been carried out for planets that transit. Extant observations divide into (1) occultation studies, (2) transit spectroscopy of varying resolution, and (3) elaborations on full-phase spectrophotometry.

A wavelength-dependent decrease in flux can be measured during *occultation* (also known as secondary eclipse), when a planet passes behind a star as seen from Earth. (Some planets, such as HD 17156 b, with its highly eccentric $e=0.68$, $\omega=112^\circ$ orbit, undergo transit, but *not* occultation.) At visible wavelengths, the occulted light from the planet consists of reflected and scattered starlight and, for hotter atmospheres, emission from the planet itself. The primary conclusion is that short-period, hot Jupiter planets tend to be *dark*. For example, Rowe et al. (2008) used the MOST telescope to find that the geometric

albedo A_g for HD 209458 is $A_g = 0.038 \pm 0.045$ in the $\lambda = 400 - 700$ nm optical band pass. More recently, Evans et al. (2013) used the STIS instrument on HST to obtain an optical and near-UV spectrum (essentially a reflection spectrum) for HD 189733b. They measured a geometric albedo $A_g = 0.40 \pm 0.12$ in the $\lambda = 290$–450 nm region and $A_g < 0.12$ for $\lambda = 290 - 450$ nm (consistent with the Rowe et al. result for HD 208458b). The planet's apparent upturn in reflectivity at short wavelengths is suggestive of a rich blue visual appearance and is interpreted to be formed by starlight reflecting from clouds coupled with high-altitude sodium absorption of red light within the overlying clear air.

At near-infrared and mid-infrared wavelengths, in the Rayleigh–Jeans tail of the star's emission spectrum, a typical hot Jupiter has a flux ratio $F_p/F_\star \gtrsim 10^{-3}$, and Spitzer has proved to be very well suited to detecting exoplanets in occultation. During the cryogenic phase of its mission, Spitzer's IRAC, IRS, and MIPS instruments were used to observe 15 exoplanets in occultation, starting with MIPS detection of HD 209458 b at 24 μm (Deming et al., 2005) and the IRAC detection of TrES-1 at 4.5 and 8 μm (Charbonneau et al., 2005). Following depletion of its liquid helium, Spitzer has continued to observe occultations with IRAC at 3.6 and 4.5 μm, and well over 30 planets now have eclipse depths measured in both of these bands.

Photometric occultation measurements at a range of different band passes amounts to a low-resolution dayside emission spectrum of an extrasolar planet, and a handful of planets have been measured in five or more bands, including J (1.2 μm), H (1.6 μm), and K (2.2 μm) from the ground, for example, Croll et al. (2011b), in addition to the Spitzer channels. For the most favorably detected transiting planets, such as HD 189733b, Spitzer's IRS instrument was used to obtain mid-infrared dayside emission spectra that were resolved into several dozen spectral resolution elements in the 5–14 μm range, with each spectral element having S/N ~ 10 (e.g., Grillmair et al., 2008). Also, for HD 189733b, HST's NICMOS has been used to obtain near-IR spectra presenting 18 spectral resolution elements in the 1.5–2.5 μm range, with S/N in each element also of order 10 (Swain et al., 2009a). These resolved spectra from IRS and NICMOS display significant variations that are strongly suggestive of molecular lines (including H_2O, NH_4, and CO), yet fitted multiparameter models are imperfect, suggesting a substantially incomplete picture of the radiative structure, chemistry, and dynamics of the atmosphere.

Naturally, there has been a vigorous effort to interpret the panoply of occultation results with theoretical models. For example, Fortney et al. (2008) suggested that strongly irradiated atmospheres (corresponding roughly to the half of the observed planets that receive the largest orbit-averaged fluxes) have thermal inversions caused by hardy molecules such as TiO or VO absorbing and reradiating starlight at low pressures high in the atmosphere. Knutson et al. (2010) presented evidence for an empirical correlation showing that chromospheric activity inhibits the formation of such thermal inversions, perhaps by destroying the inversion-producing molecules through an elevated flux of high-energy photons. More recently, Madhusudhan (2012) (see also Madhusudhan and Seager, 2010) had suggested that C/O ratios form an additional dimension by which planets can exhibit planet-to-planet variation in occultation depths.

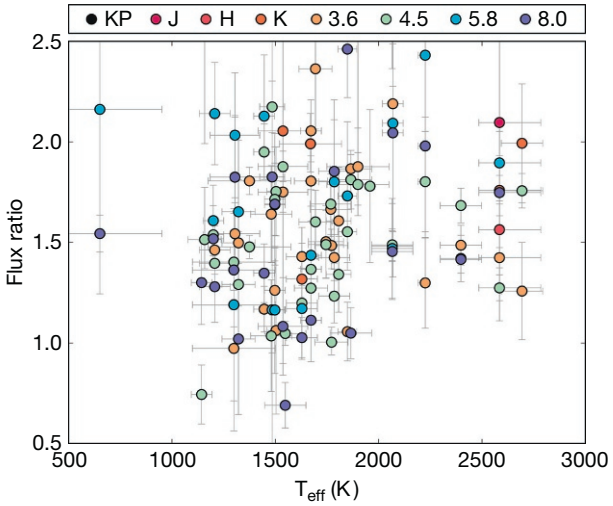


Figure 8 The current collection of broadband occultation measurements. Planets are ordered on the *x*-axis by the equilibrium temperatures, $T_{eq} = (R_\star^{1/2} T_\star)/((2a)^{1/2}(1-e^2)^{1/8})$, that they would have if they were $A_g = 0$ blackbodies reradiating from the full planetary surface area. Measured occultation depths are expressed as ratios of the observed eclipse depth in the band of interest relative to the expected depth in that band if the planet were a uniformly reradiating $A_g = 0$ blackbody. Colors are K (2.2 μm), purple; 3.6 μm, blue; 4.5 μm, green; 5.8 μm, orange; and 8 μm, red.

Figure 8 consolidates a full range of broadband occultation results onto a single plot. The planets are ordered according to specific orbit-averaged received flux, as measured by T_{eq}. Occultation depths in the various infrared bands are transformed to flux ratios, F_λ relative to the expectation for a zero albedo 4π blackbody reradiator. It is clear from the diagram that there is *little or no* systematic degree of similarity among the planets. The data are essentially consistent with random flux ratios relative to $<F_\lambda> = 1.5$, with standard deviation $\sigma_\lambda = 0.5$. This distribution of measurements suggests a complex mix of contributing factors (stemming either from conditions on the planets or from errors introduced in the measurement pipelines), with the attendant applicability of the central limit theorem. The $<F_\lambda> = 1.5$ mean value, however, implies that on the whole, exoplanets are *partially* redistributing their received flux via reradiation from the night side (for further discussion, see Langton et al. (2014)). Certainly, Figure 8 is in concordance with the view expressed by Hansen et al. (2014) that, to the precision of the measurements, the observed emission spectra of exoplanets are consistent with blackbodies.

During transit, starlight flows through the atmosphere of the planet all along the day–night terminator. A stellar spectrum obtained during transit thus contains an admixture of the planet's transmission spectrum, which holds out the possibility of obtaining a detailed probe of the atmospheric column. An early application of transit spectroscopy is found in Charbonneau et al. (2002), who used HST to detect sodium in the atmosphere of HD 209458b by noting that the transit depth associated with a band pass covering the sodium resonance doublet at 589.3 nm was deeper than in adjacent bands. HST has also been used to detect strong evidence of water vapor, methane, and carbon monoxide (e.g., Swain et al.,

2009a,b) in the 1.5–2.5 μm transit spectra of HD 209458b and HD 189733b.

For planets transiting $V > 8$ primaries, a variety of broadband transmission spectra have been obtained for a variety of planets. Interpretation of these spectra is challenging, and, in parallel with the situation for occultations, firm conclusions are hampered by intrinsically low spectral resolution, low signal-to-noise ratio, and the need to adopt a whole range of model assumptions. For more detail, the reader is advised to consult the review by Tinetti et al. (2013). JWST is expected to have significant utility for transit spectroscopy and will be able to usefully probe the atmospheres of planets transiting M-dwarf stars (Deming et al., 2009).

There is an exciting possibility of directly detecting planetary surface winds by observing the Doppler shift of spectral absorption lines that are created by the starlight shining through moving parcels of air at the planet's day–night terminator. Snellen et al. (2010) used the VLT to obtain a high-resolution transit spectroscopy of HD 209458b with $R \sim 100{,}000$ at wavelengths near $\lambda = 2$ μm. By cross correlating on dozens of expected CO lines, these authors were able to produce a time-dependent measurement of the average velocity of the atmosphere at the planetary terminator during the course of the transit. With the orbital motion subtracted out, Snellen et al. (2010) measured a net blueshift of 2 ± 1 km s^{-1} and suggested that this represents a global flow from the dayside hemisphere to the nightside hemisphere at atmospheric levels in the range 0.01 bar $< P <$ 0.1 bar. Showman et al. (2013) found that generic flow patterns with this large net blueshift can be created within current extrasolar planet atmospheric circulation simulations (see, e.g., Showman et al., 2009), by varying the amount of frictional drag in the atmosphere.

For planets in the 'ungiant' regime, atmospheric measurements obtained via transit spectroscopy can potentially help to sort out the compositional degeneracies that prevent strong conclusions being drawn from the measurements of bulk planetary density (e.g., those shown in **Figure 5**). A planet with a refractory core covered by a deep solar-composition atmosphere will have a wavelength-dependent transit depth that can be potentially distinguished, using current observational sensitivity, from the signature produced by a planet with an atmosphere (and an overall structure) dominated by high-molecular-weight species such as water. For example, if the atmosphere of a planet such as Gliese 1214b, with $M_p - 6.6 M_\oplus$, $\rho_p = 0.35 \rho_\oplus$ and $T_p \sim 500$ K (Charbonneau et al., 2009), is primarily composed of H_2, then the resulting atmospheric scale height, $H = k_B T/\mu g$, will be of order $H \sim 200$ km. Given that Gliese 1214b has an M4.5 V primary, this implies that $H/R_\star \sim 0.0014$. For an H to $R_\star$ ratio of this magnitude, ground-based measurements are presently capable of distinguishing between a clear H_2-dominated solar-composition atmosphere, which will show measurable differences in planetary radius as a function of wavelength, $\Delta(R_p/R_\star) \sim 0.005$ (de Mooij et al., 2012), and a high-molecular-weight atmosphere with a small-scale height and a transmission spectrum that presents much smaller variations.

Measurements of $\Delta(R_p/R_\star)$ taken at different band passes during transits by Gliese 1214b have indicated that the transmission spectrum is fairly flat between $\lambda = 0.6$ μm and $\lambda = 5$ μm (Bean et al., 2010, 2011; Berta et al., 2012; Croll et al., 2011a; de Mooij et al., 2012; Désert et al., 2011). As of early 2014, a number of these measurements are in conflict, however, and while a clear solar-composition atmosphere is inconsistent with the data, Gliese 1214b could be presenting either a hazy low-metallicity H–He-dominated atmosphere or a low-scale height atmosphere composed of species (such as H_2O) of high molecular weight. Very recent work by Kreidberg et al. (2013) reports intensive HST observations of Gliese 1214b (15 transits over 60 telescope orbits) that are sensitive enough to reveal absorption features from constituents such as H_2O, CO_2, and NH_4 in a high-molecular-weight atmosphere. A flat transmission spectrum between 1.1 and 1.7 μm was observed, suggesting that Gliese 1214b's atmosphere *must* contain clouds or high-altitude hazes. In addition, other planets with $P < 100$ days and $M_p < 30 M_\oplus$ are nearly certain to be discovered in transit across bright low-mass primaries. For example, Bonfils et al. (2012) had announced the detection of Gliese 3470b, which has $M_p = 14 M_\oplus$, $R_p = 4.2 R_\oplus$, and $T_p \sim 800$ K and which orbits an $R_\star \sim 0.5 R_\odot$ M1.5 V primary. Spectrographic observations of Gliese 3470b by Crossfield et al. (2013) using the Keck telescope are consistent with a flat spectrum between 2.09 and 2.36 μm, which rule out a cloud-free atmosphere in chemical equilibrium and which are again suggestive of high-altitude haze or clouds. The prospects for improved characterization of planets in the 'ungiant' category are especially bright, given that NASA's forthcoming TESS mission will locate a substantial fraction of the transiting short-period planets with radii $R_p \gtrsim 2.0 R_\oplus$ orbiting stars with $V < 12$.

Transit and occultation measurements act as effective bookends to *full-phase* photometric measurements, in which planetary emission is monitored through the course of an orbit. The most frequently cited example of this type of observation was published by Knutson et al. (2007), who presented a high-quality 8 μm light curve for HD 189733b that spans the 29 h interval from just before the planet's primary transit to just after its secondary eclipse. Adopting the assumption that the planet is in a spin-synchronous rotation with its orbit, Knutson et al. (2007) derived a global temperature distribution for the planetary photosphere as a function of longitude and found that the hottest region of the planet is shifted east of the substellar point (see Cowan and Agol (2008) for the details of the mapping technique). The longitudinal shift of the planet's temperature maximum is attributed to superrotating $v \sim 1$ km s^{-1} equatorial winds, which are a robust feature that emerge from a wide variety of hydrodynamic models of the surface flows on tidally locked close-in extrasolar planets (see, e.g., Showman and Polvani, 2011; Showman et al., 2009). In the past several years, a number of additional infrared phase curves in assorted band passes have been produced for a variety of strongly irradiated planets. Examples include Knutson et al. (2009b)'s follow-on partial-orbit MIPS 24 μm and Knutson et al. (2012)'s full-orbit 3.6 and 3.5 μm IRAC phase curves for HD 189733 (which corroborate the basic picture inferred by Knutson et al., 2007) and a full-orbit optical phase curve from the Kepler satellite for HAT-P-7b (Borucki et al., 2009) that shows evidence of both dayside thermal emission and reflected light. Full- or partial-orbit near-IR phase curves have also been published for HD 149026b (Knutson et al., 2009a), WASP-18b (Maxted et al., 2013), and WASP-12b (Cowan et al., 2012), all of which are hot, short-period planets on nearly

circular orbits. The phase curves for these systems suggest that the selection of planets surveyed have a large range of efficiency in their ability to equilibrate day–night temperature distributions.

Figure 9 shows phase curves in several band passes for the eccentric planet HAT-P-2b, as presented by Lewis et al. (2013). The location of the substellar point (high noon) must move on planets with eccentric orbits (if the rotation period equals the orbital period, rotation is referred to as synchronous, e.g., the Moon) and hence, their atmospheres are virtually certain to display strong time-dependent meteorology. Phase-curve observations of such planets can potentially determine atmospheric radiative time constants, tidal luminosities, and pseudosynchronous rotation rates, as well as the global manifestation of time-dependent storms (e.g., Langton et al., 2014).

10.21.5 Planet Formation

The vast majority of work on planetary formation has been done to explain the properties of our solar system, for which we have the most detailed data. We begin with a synopsis of these models and then discuss some recent models of the formation and migration of exoplanets that travel on short-period orbits.

10.21.5.1 An Overview Informed by the Solar System

Models of the solar system formation are based upon data from our own solar system and observations of the circumstellar environments of young stars. The nearly planar and almost circular orbits of the planets in our solar system argue strongly for planetary formation within flattened circumstellar disks. When a molecular cloud core undergoes gravitational collapse, an occurrence triggered by loss of magnetic field and turbulent velocity support, its spin angular momentum leads to the formation of pressure-supported protostars surrounded by rotationally supported disks, for example, Terebey et al. (1984). Molecular cloud cores are subcondensations of mass, $M \gtrsim M_{\odot}$, characteristic size $r \sim 0.25$ pc, and rotation rate typically $\sim 10^{-13}$ Hz that are found within giant molecular cloud complexes in the galactic disk that typically contain 10^4 to $10^5 M_{\odot}$.

Circumstellar disks are commonly observed in association with newly formed stars (Mamajek, 2009) and are analogous to the primordial solar nebula that was initially conceived by Kant and Laplace to explain the observed properties of our

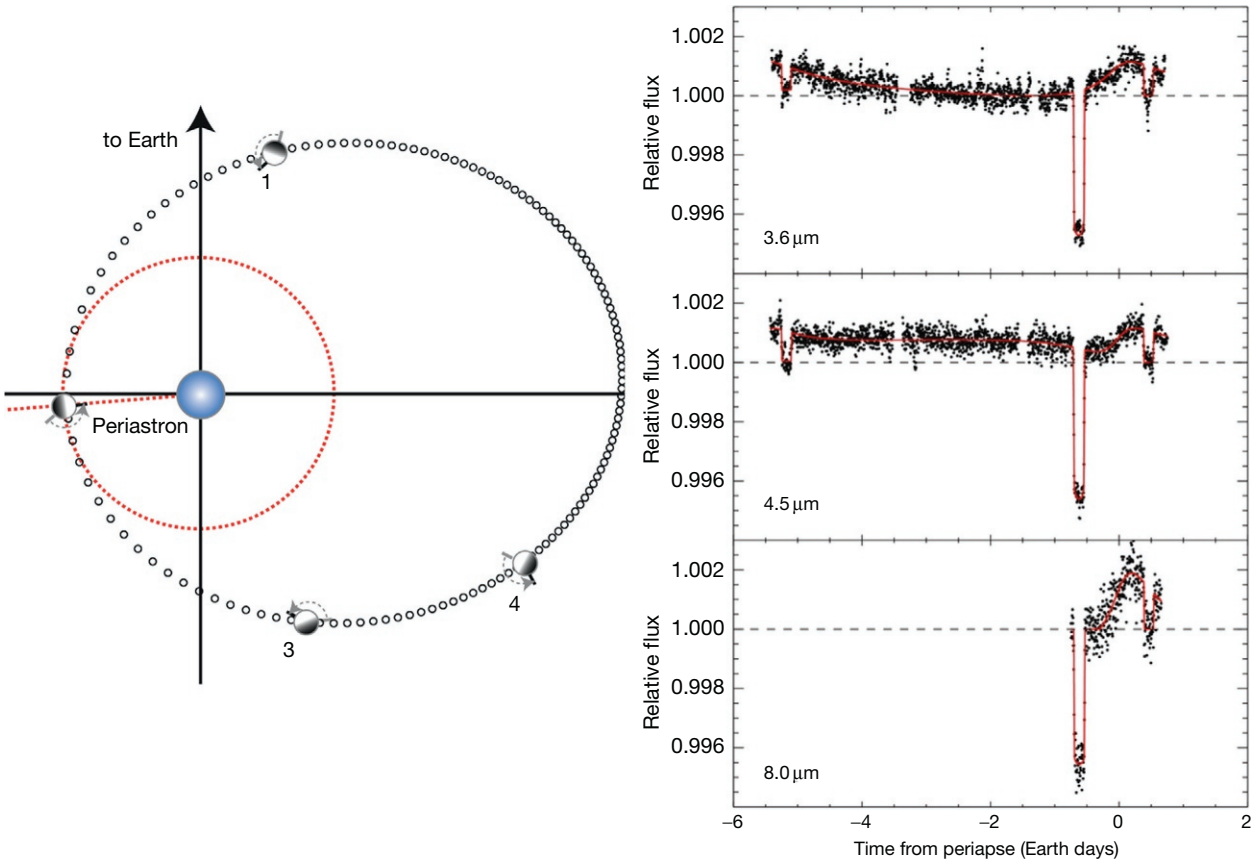


Figure 9 Photometric IR phase curves for the eccentric planet HAT-P-2b, with orbital period, $P = 5.633$d; eccentricity, $e = 0.511$; and $M_p = 9M_J$ (adapted from Lewis NK, Knutson HA, Showman AP, et al. (2013) *Astrophysical Journal* 766: 95). The left panel shows the geometry of the HAT-P-2b orbit, with periastron passage occurring 14 h after the primary transit. The panels on the right display the light curves for the 3.6 (top), 4.5 μm (middle), and 8.0 μm photometry. Optimized phase, transit, and secondary eclipse fits are overplotted as red line (see Lewis et al. (2013) for details). The dashed line represents the stellar flux level.

solar system. For a more detailed historical background, see Woolfson (1993). Planets form within such disks and play a major role in the later stages of disk evolution. Terrestrial planets grow via pairwise accretion of solid bodies. In the generally favored core accretion paradigm, giant planets begin their growth as do terrestrial planets, but they become massive enough to accumulate substantial amounts of gas before the protoplanetary disk dissipates. Under some circumstances, as described, for example, by Gammie (2001), giant planets may form directly from the protostellar disk via gravitational instability (see, e.g., Boley, 2009; Durisen et al., 2007), especially at large ($d \gtrsim 40$AU) from the parent star.

Within the solar system, bodies up to the mass of the Earth consist almost entirely of condensable (under reasonable protoplanetary disk conditions) material, and the fraction of highly volatile gas increases with mass above that of the Earth. The most massive planets contain a considerable fraction of light gases. About 90% of Jupiter's mass is H and He, and these two light elements make up $\sim$75% of Saturn. In both cases, these planets contain $\sim 30 M_{\oplus}$ in elements other than hydrogen and helium. The large amounts of H and He contained in Jupiter and Saturn imply that these planets must have formed within $\sim 10^7$ years of the collapse of the solar system's natal cloud, before the gas in the protoplanetary disk was swept away. The two largest planets in our solar system are generally referred to as *gas giants*, even though these elements are not gases at the high pressures that most of the material in Jupiter and Saturn is subjected to. Analogously, Uranus and Neptune are frequently referred to as *ice giants*, even though the astrophysical 'ices' such as H_2O, CH_4, H_2S, and NH_3 that models suggest make up the majority of their mass (Hubbard et al., 1995) are in fluid rather than solid form. Note that whereas H and He *must* make up the bulk of Jupiter and Saturn because no other elements can have such low densities at plausible temperatures, it is possible that Uranus and Neptune are primarily composed of a mixture of 'rock' and H/He, although the high observed CH_4/H_2 ratio in their atmospheres (30–40 $\times$ the solar value) makes this somewhat unlikely. This ambiguity in basic structure also holds for the majority of the ungiant exoplanets whose mass and radius have both been measured.

Lithium and heavier elements constitute $< 2\%$ of the mass of a solar-composition mixture. The *atmospheric* abundances of volatile gases heavier than helium (excluding neon, which is thought to be substantially depleted through gravitationally induced settling) are $\sim 3 \times$ solar in Jupiter (Young, 2003), a bit more enriched in Saturn, and substantially more for Uranus and Neptune. The *bulk* enhancements in heavy elements relative to the solar value are roughly 5, 15, and 300 times for Jupiter, Saturn, and Uranus/Neptune, respectively. Thus, all four giant planets accreted solid material substantially more effectively than gas from the surrounding nebula. Moreover, the total mass in heavy elements varies by only a factor of a few between the four planets, while the mass of H and He varies by about two orders of magnitude between Jupiter and Uranus/Neptune.

There is a vast observational literature that details the presence, the properties, and the evolution of disks of the solar system dimensions around pre-main-sequence stars (Williams and Cieza, 2011). The existence of disks on scales of a few tens of astronomical units is inferred from the power law spectral energy distribution in the infrared over more than two orders of magnitude in wavelength (Andrews et al., 2010; Chiang and Goldreich, 1997; Haisch et al., 2001; Megeath et al., 2004). Observations of infrared excesses in the spectra of young stars suggest that the lifetimes of protoplanetary disks span the range of 10^6–10^7 years (Alencar and Batalha, 2002; Ercolano and Koepferl, 2012; Strom et al., 1993). Much lower-mass second-generation debris disks composed of dust eroded from larger solid objects that accreted within protoplanetary disks are observed around main-sequence stars, with younger stars typically (but far from always) possessing more massive debris disks (Krist et al., 2012; Meyer et al., 2004). Many debris disks appear to have inner holes, and lower-mass stars typically retain their debris disks for more time (Wyatt, 2008).

Although to a first approximation the gas in the disk is centrifugally supported in balance with the star's gravity, negative radial pressure gradients provide a small outwardly directed force that acts to reduce the effective gravity, so the gas rotates about the star at slightly less than the Keplerian velocity. Small solid bodies (dust grains) rotate with the gas. Large solid bodies orbit at the Keplerian velocity, and medium-sized particles move at a rate intermediate between the gas velocity and the Keplerian velocity; thus medium-sized and larger bodies are subjected to a headwind from the gas (Adachi et al., 1976). This headwind removes angular momentum from the particles, causing them to spiral inward toward the star. This inward drift can be very rapid, especially for particles whose coupling time to the gas is similar to their orbital period; smaller particles drift less rapidly because the headwind they face is not as strong, whereas large particles drift less because they have a greater mass to surface area ratio. Orbital decay times for meter-sized particles at 1 AU from the Sun have been estimated to be only $\sim$100 years (Weidenschilling, 1977). The large radial velocities of bodies in this size range relative to both larger and smaller particles imply frequent collisions, so most solid bodies may pass through the critical size range quickly without substantial radial drift. However, it is also possible that a large amount of solid planetary material is lost from the disk in this manner. Solid bodies larger than $\sim$1 km in size face a headwind only slightly faster than meter-sized objects, and because of their much greater mass to surface area ratio, they suffer far less orbital decay from interactions with the gas in their path. Thus, kilometer-sized or larger planetesimals appear to be reasonably safe from loss (unless they are ground down to smaller sizes via disruptive collisions) until some of these planetesimals grow into planetary-sized bodies.

The initial composition of the protoplanetary disk is assumed to still be essentially identical to that of the protosun. Dust within a protoplanetary disk initially agglomerates via sticking/local electromagnetic forces that produce inelastic collisions. Particles gradually settle toward the disk midplane as they grow.

The conceptual bottleneck created by the short residence times of meter-sized particles in protoplanetary disks is perhaps the least well-understood phase of solid planet formation. It is not yet clear how agglomeration from cm-sized pebbles to km-sized or larger bodies that are referred to as planetesimals occurs. Collective gravitational instabilities (Goldreich and Ward, 1973; Safronov, 1969) might be important, although turbulence may prevent protoplanetary dust layers from becoming thin enough to be gravitationally unstable (Weidenschilling and Cuzzi, 1993). The interactions

between small particles and gas in a rotating disk involve complicated fluid-particle interactions, and planetesimal formation is a very active research area. Recent work suggests that the answer may lie in the joint action of aerodynamic and gravitational instabilities to agglomerate particles (see the review of Chiang and Youdin (2010)). Some models suggest that pebble-sized particles accumulated into planetesimals more than 100 km in size (Johansen et al., 2012).

The primary perturbations on the Keplerian orbits of kilometer-sized and larger bodies in protoplanetary disks are pairwise gravitational interactions and physical collisions (Safronov, 1969). These interactions lead to accretion (and in some cases erosion and fragmentation) of planetesimals. Gravitational encounters are able to stir planetesimal random velocities up to the escape speed from the largest common planetesimals in the swarm (Safronov, 1969). The most massive planets have the largest gravitationally enhanced collision cross sections and accrete almost everything with which they collide. If the random velocities of most planetesimals remain much smaller than the escape speed from the largest bodies, then these 'planetary embryos' grow extremely rapidly (Greenzweig and Lissauer, 1992). The size distribution of solid bodies becomes quite skewed, with a few large bodies growing much faster than the rest of the swarm in a process referred to as runaway accretion (Greenberg et al., 1978; Wetherill and Stewart, 1989). If embryos dominate the stirring, then $\dot{M} \propto M^{2/3}$; in this circumstance, if individual embryos control the velocity in their own zones, larger embryos take longer to double in mass than do smaller ones, although embryos of all masses continue their runaway growth relative to surrounding planetesimals; this phase of rapid accretion of planetary embryos is known as *oligarchic growth* (Kokubo and Ida, 1998). Eventually, planetary embryos accrete most of the (slowly moving) solids within their gravitational reach, and the runaway/oligarchic growth phase ends.

The self-limiting nature of runaway and oligarchic growth implies that massive planetary embryos form at regular intervals in semimajor axis. The agglomeration of these embryos into a small number of widely spaced terrestrial planets necessarily requires a stage characterized by large orbital eccentricities, significant radial mixing, and giant impacts. At the end of the rapid-growth phase, most of the original mass is contained in the large bodies, so their random velocities are no longer strongly damped by energy equipartition with the smaller planetesimals. Mutual gravitational scattering can pump up the relative velocities of the planetary embryos to values comparable to the surface escape velocity of the largest embryos, which is sufficient to ensure their mutual accumulation into planets. The large velocities imply small collision cross sections and hence long accretion times.

Once the planetary embryos have perturbed one another into crossing orbits, their subsequent orbital evolution is governed by close gravitational encounters and violent, highly inelastic collisions. Terrestrial planets continue to grow by pairwise accretion of solid bodies until the spacing of planetary orbits becomes sufficient for the configuration to be stable to gravitational interactions among the planets for the lifetime of the system (Chambers, 2001; Hansen and Murray, 2012; Laskar, 2000; Lissauer, 1995; Safronov, 1969). This process has been studied using N-body integrations of planetary embryo orbits. Many of the simulations of this type have endeavored to reproduce our solar system; they generally begin with about $2M_{\oplus}$ of material spread through the terrestrial planet zone, typically divided (not necessarily equally) among hundreds or thousands of bodies. These simulations generally display a high efficiency of converting seed embryos into terrestrial planets, and the end result is generally the formation of 2–5 terrestrial planets on a timescale of about 10^8 years. Some of these systems look startlingly similar to our inner solar system, with the caveat that the emergent planets often travel on more eccentric orbits than those of the terrestrial planets, although the simulations that begin with the largest numbers of bodies yield eccentricities more similar to those in our solar system (O'Brien et al., 2006). It is possible that the solar system is by chance near the quiescent end of the distribution of terrestrial planets, but more likely, a damping process of some type plays a role. Processes such as fragmentation and gravitational interactions with a remaining population of small debris and interactions with a remnant gas disk (Ogihara et al., 2007) have been invoked to lower the characteristic eccentricities and inclinations of the ensemble of terrestrial planets.

An important result of these N-body simulations is that planetary embryo orbits execute a random walk in semimajor axis as a consequence of successive close encounters. The resulting widespread mixing of material throughout the terrestrial planet region diminishes any chemical gradients that may have existed when planetesimals formed, although some correlations between the final heliocentric distance of a planet and the region where most of its constituents originated are preserved in the simulations. Nonetheless, these dynamical studies imply that Mercury's high iron abundance is unlikely to have arisen from chemical fractionation in the solar nebula.

The mutual accumulation of numerous planetary embryos into a small number of planets must have entailed many collisions between protoplanets of comparable size. Mercury's original silicate mantle was probably partially stripped off by one or more of such giant impacts, leaving behind an iron-rich core. Accretion simulations also lend support to the giant impact hypothesis for the origin of the Earth's Moon; during the final stage of accumulation, an Earth-sized planet is typically found to collide with several objects as large as the Moon and frequently one body as massive as Mars (Chambers, 2001).

Models for the formation of gas giant planets were reviewed by Wuchterl et al. (2000) and Helled et al. (2013). Starlike direct quasispherical collapse is not considered viable, because of both the observed brown dwarf desert mentioned in the preceding text and theoretical arguments against the formation of Jupiter-mass objects via fragmentation (Bodenheimer et al., 2000). The theory of giant planet formation that is favored by most researchers is the *core-nucleated accretion model*, in which the planet's initial growth resembles that of a terrestrial planet but eventually becomes sufficiently massive (several $M_{\oplus}$) that it is able to accumulate substantial amounts of gas from the surrounding protoplanetary disk. The only other hypothesis receiving significant attention is the *gas instability model*, in which the giant planet forms directly from the contraction of a clump that was produced via a gravitational instability in the protoplanetary disk (Durisen et al., 2007). Formation of giant planets via gas instability has never been demonstrated for realistic disk conditions. Moreover, this model has difficulty explaining the supersolar abundances of heavy

elements in Jupiter and Saturn, and it does not explain the origin of planets like Uranus and Neptune. Nonetheless, it is possible that some giant planets form via disk instability.

The core-nucleated accretion model (Mizuno, 1980; Pollack et al., 1996) has been reviewed by Lissauer and Stevenson (2007), with further details given in D'Angelo et al. (2011). The model relies on a combination of planetesimal accretion and gravitational accumulation of gas. In this theory, the core of the giant planet forms first by accretion of planetesimals, while only a small amount of gas is accreted. Core accretion rates depend upon the surface mass density of solids in the disk and physical assumptions regarding gas drag, etc. (Inaba et al., 2003; Lissauer, 1987). The escape velocity from a planetary embryo with $M > 0.1M_{\oplus}$ is larger than the sound speed in the gaseous protoplanetary disk. Such a growing planetary core first attains a quasistatic atmosphere that undergoes Kelvin–Helmholtz contraction as the energy released by the accretion of planetesimals and gas is radiated away at the photosphere.

During the runaway planetesimal accretion epoch, the protoplanet's mass increases rapidly. The internal temperature and thermal pressure increase as well, preventing substantial amounts of nebular gas from falling onto the protoplanet. When the planetesimal accretion rate decreases, gas falls onto the protoplanet more rapidly. The protoplanet accumulates gas at a gradually increasing rate until its gas mass is comparable to its heavy element mass. The key factor limiting gas accumulation during this phase of growth is the protoplanet's ability to radiate away energy and contract (Hubickyj et al., 2005; Pollack et al., 1996). The rate of gas accretion then accelerates more rapidly, and a gas runaway occurs. The gas runaway continues as long as there is gas in the vicinity of the protoplanet's orbit. The protoplanet may cut off its own supply of gas by gravitationally clearing a gap within the disk (Lin and Papaloizou, 1979). Lissauer et al. (2009) used a three-dimensional adaptive mesh refinement code (D'Angelo et al., 2002, 2003) to follow the flow of gas onto an accreting giant planet, allowing the determination of final planetary mass as a function of the time-varying properties (density, temperature, viscosity, longevity, etc.) of the surrounding disk. Such a self-regulated growth limit provides a possible explanation to the observed mass distribution of extrasolar giant planets.

Core-nucleated accretion also offers an explanation for the formation of low-density ungiant exoplanets. These bodies could be cores that grew massive enough to accrete some H_2 and He from their planetary disks but were unable to accrete enough light gases to become giants (Rogers et al., 2011).

10.21.5.2 How Did Short-Period Planets Arrive at their Observed Locations?

Quite aside from the zeroth-order questions related to their atmospheric properties and internal structures, the origins of the short-period planets are not fully explained. It is clear from **Figure 2** that the hot Jupiters represent a population distinct from the short-period ungiants, and so it is reasonable to consider their formation scenarios separately.

It is generally assumed that hot Jupiters were formed through the core accretion process, with assembly occurring at distances at least a few AU from their host stars, where more solids, especially water ice, were available. As discussed in the preceding text, in this standard core accretion scenario for giant planet formation, a protoplanetary core composed largely of ices and rock reaches a mass of $\sim 10M_{\oplus}$, at which point it begins to rapidly accrete gas from the surrounding nebular disk. Thereafter, as a result of dynamical or hydrodynamic processes, the planet is delivered into a short-period orbit. It is worth noting, however, that while disfavored, *in situ* formation for hot Jupiters has not been strictly ruled out. Detailed calculations by Bodenheimer et al. (2000), for example, suggest that hot Jupiters can form in place under certain conditions.

The 'classical' formation scenario for hot Jupiters was outlined by Lin et al. (1996), shortly after the discovery of 51 Peg b. In this picture, the planet grows to near its original mass at a distance several AU from the star. A planet with 51 Peg b's mass is large enough to maintain a well-cleared annular gap in the disk that brackets the planet's radial location. Once formed, the planet spirals inward via the process of type II migration (Lin and Papaloizou, 1986) on the parent protoplanetary disk's viscous timescale, $\tau_{\nu} = r_{\mathrm{d}}^{1/2}\nu \sim 10^6$ years. The final location is controlled either by disk truncation at an inner magnetospheric cavity or via tidal interaction with the young, rapidly spinning parent star. This basic picture has been elaborated substantially during the past two decades. For recent overviews, see Lubow and Ida (2011) and Kley and Nelson (2012).

The disk migration scenario avoids the serious problems associated with *in situ* formation, but is not itself without problems. One difficulty lies with the so-called type I migration (Ward, 1997), which is thought to affect planets that have insufficient mass to clear and maintain a gap in the protostellar disk. The timescale for type I migration $\tau_1 \sim (M_{\mathrm{p}}/M_{\star})$ $c_{\mathrm{s}}\Omega \sim (M_{\mathrm{p}}/M_{\oplus})10^5$ years is substantially shorter than the timescale that appears to be required to assemble a core large enough to initiate rapid gas accretion (Movshovitz et al., 2010). The mismatch hints at the possibility that a fundamental process is either missing or is not fully understood. A further problem arises from observed statistics of systems with massive, short-period planets. Hot Jupiters appear almost exclusively in single-planet configurations (Steffen et al., 2012) (with the υ Andromedae planets providing a notable exception).

Purely dynamical mechanisms have received increased attention in connection with the production of hot Jupiters. An early and influential study was carried out by Rasio and Ford (1996), who attributed hot Jupiters to the outcome of catastrophic planet–planet scattering. In their picture, extant short-period planets are survivors of strong gravitational encounters, in which one planet was either ejected or placed in a long-period orbit and the other is left in a high-eccentricity orbit with modest semimajor axis. Given a sufficiently close periastron distance, a high-eccentricity planet will subsequently undergo orbital circularization.

The process of Lidov–Kozai migration with tidal friction has grown rapidly in popularity during the last several years. In the Lidov–Kozai scenario (see, e.g., Eggleton and Kiseleva-Eggleton, 2001; Fabrycky and Tremaine, 2007; Wu and Murray, 2003), a planet destined to become a hot Jupiter forms at a distance of several AU from its parent star in an orbit that is substantially inclined at angle I_{b} to the orbit of a distant binary companion of mass $M_{\mathrm{b}\star}$. If the mutual orbital inclination between the stellar binary companion is larger or,

better, substantially larger than the Lidov–Kozai critical angle, $I_c = \arccos[(3/5)^{1/2}] \sim 40^\circ$ (assuming $M_{b\star} \gg M_p$ and an initially circular orbit for b), the planet is compelled to undergo *Lidov–Kozai cycles* in which the Jacobi energy $E_J = (1-e^2)^{1/2}\cos I$ is conserved, in which the planet experiences periodic cycling between successive states of high orbital eccentricity and high mutual inclination. The timescale for these cycles is

$$\tau_{LK} = \frac{2P_C^2}{3\pi P_B}\frac{M_\star + M_B + M_c}{M_c}\left(1-e_c^2\right)^{3/2} \quad [12]$$

For large initial I_B, the planet's periastron distance during the high-eccentricity phase of the cycle can be small enough so that dynamical tides acting on the planet are capable of efficiently turning orbital energy into heat, thereby shrinking the semimajor axis, a, of the planetary orbit. The Lidov–Kozai cycles are eventually destroyed when the general relativistic precession rate

$$\dot{\omega}_{GR} = \frac{3G^{3/2}(M_\star + M_B)^{3/2}}{a_B^{5/2}c^2(1-e_B^2)} \quad [13]$$

exceeds the Lidov–Kozai cycling timescale, that is, $2\pi/\dot{\omega}_{GR} < \tau_{LK}$. At this point, the planet shifts to a secular evolution characterized by steady orbital decay and circularization, with the end product being a hot Jupiter.

The Lidov–Kozai process delivers planets into orbits in which the planetary orbital angular momentum and the stellar spin angular momentum are initially largely uncorrelated.

A slew of measurements of the sky-projected angle, λ, between the stellar rotation and planetary orbital angular momentum vectors using the Rossiter–McLaughlin spectroscopic technique (McLaughlin, 1924; Rossiter, 1924) indicate that a substantial fraction of hot Jupiter orbits are badly misaligned with the equators of their parent stars.

10.21.6 Questions for the Next 10 Years and Beyond

There are few precedents in the history of science in which a discipline moves so rapidly from shaky disrepute through a golden age of discovery and into a mature field of inquiry. In less than two decades, the study of extrasolar planets has accomplished all of this – answering questions that were posed at the dawn of the scientific era, while affording tantalizing glimpses of revelations to come. This rapid progress partially obscures the fact that the extrasolar planets are *fundamentally alien*. Virtually none of their properties, either statistical or physical (aside from the scarcity of planets with $30M_\oplus < M_p < 100M_\oplus$), were predicted or anticipated (Lissauer, 1995, 2006), and theorists struggle to understand the zeroth-order features of the planetary distribution. At the current moment, it is thus useful to conclude by listing some of the most important unsolved questions.

- *Why is our inner solar system empty*? The *Kepler* mission and the high-precision Doppler velocity surveys have shown that the default mode of planet formation generates multiple super-Earth-mass planets with orbital periods $P < 100$ days. Our own solar system, however, contains nothing interior to Mercury's 88-day orbit. Furthermore, given the *presence* of Jupiter, with a mass and period combination that appears to be at least somewhat unusual, is there a possible connection between these two abnormal solar system attributes? Recent work of Walsh et al. (2011) suggests that both Jupiter and Saturn experienced significant radial migration, and the so-called Nice model (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005) posits that substantial migration of the outer planets occurred during the solar system's formation phases. A provocative, and as yet unanswered, question is whether such a process – generated, for example, via sweeping secular resonance – could have thwarted planet formation in our inner solar system.
- *Where in the protostellar disks did the observed planets form*? Our solar system has long been thought to be the product of planet formation that occurred more or less *in situ*, via core accretion for the giant planets and via planetesimal accretion for the terrestrial planets. The detection of 51 Peg b, along with other early extrasolar planet discoveries, led to the suggestion, which soon became a paradigm, that long-distance planetary migration can and does occur. At present, there are substantial arguments both in favor and against *in situ* formation of close-in ungiants. For hot Jupiters, where initial formation at large astrocentric distance seems more probable, it is not clear whether the Lidov-Kozai Cycles with Tidal Friction mechanism or quiescent long-distance type II disk migration is the dominant process.
- *What are the bulk compositions of the ungiants*? If one considers $\log_{10}M$, the most prominent mass gaps among all the bodies populating the solar system are those between Jupiter and the Sun, with $\log_{10}M_\odot/M_J = 3.02$, and between Earth and Uranus, with $\log_{10}(M_{Ura}/M_\oplus) = 1.16$. While it has long been possible to explore the properties of objects in the Sun–Jupiter gap through observations of low-mass stars and, more recently, brown dwarfs (Chabrier and Baraffe, 2000), the mere presence of extrasolar planets with $M_\oplus < M_p < M_{Ura}$ orbiting main-sequence stars has been known only since the discovery of Gliese 876 d, with $M_p = 7.5M_\oplus$ (Rivera et al., 2005), and the first density estimate became available only with the unveiling of Gliese 1214b, which has $M_p = 6.3M_\oplus$, $R_p = 2.7R_\oplus$, and $\rho_p = 1.67\,g\,cm^{-3}$ (Charbonneau et al., 2009). Structural models, as calculated by, for example, Rogers and Seager (2010), can easily reproduce Gliese 1214b's mass and radius with configurations drawn from three board classes of models, (i) rock and iron planets with massive H/He or H_2 atmospheres, (ii) rock–ice cores surrounded by H–He envelopes of solar nebular composition, and (iii) the so-called water worlds composed primarily of H_2O.

As we have described, attempts to characterize the atmosphere of Gliese 1214b have been met with ambiguity. It is likely that only when high-quality transmission spectra and interior mass distribution measurements are obtained will a clear understanding of the true range of compositions and structures of the ungiants emerge.

- *What (if any) is the observable role of magnetohydrodynamics*? The Earth and the four larger planets in our solar system all

harbor intrinsic magnetic fields that are generated by dynamo action in the planetary interiors. It is therefore natural to expect that many, perhaps most, short-period extrasolar planets with $M_p >$ several $M_\oplus$ also have magnetic fields, and to first approximation, these fields can be estimated by assuming that the *Elsasser number*, $\Lambda = \sigma_i B^2/2\rho\Omega$, is roughly constant across the aggregate of transiting planets ranging from Earth-like to Jupiter-like worlds. If we further assume synchronous rotation (which should hold for hot Jupiters on circular orbits), then $1/\Omega \propto a^{3/2}$. The upper atmospheres of planets with $P \lesssim 5$ days are hot enough ($T \gtrsim 1200$ K) that alkali metals (e.g., Na and K) have ionization fractions that are sufficient to couple the planetary magnetic field to the atmospheric hydrodynamic flow (see, e.g., Shu, 1992). Studies, such as those by Batygin and Stevenson (2010), suggest that ohmic dissipation is responsible for heating planetary interiors, thereby explaining much of the observed radius anomaly trend. Do MHD-mediated influences have a zeroth-order effect on exoplanet atmospheres and can some of the bewildering variation in atmospheric emission be understood as arising from magnetohydrodynamic phenomena?

References

Adachi I, Hayashi C, and Nakazawa K (1976) *Progress of Theoretical Physics* 56: 1756.
Adams ER, Seager S, and Elkins-Tanton L (2008) *Astrophysical Journal* 673: 1160.
Alcock C, Allsman RA, Alves D, et al. (1997) *Astrophysical Journal* 486: 697.
Alencar SHP and Batalha C (2002) *Astrophysical Journal* 571: 378.
Andrews SM, Wilner DJ, Hughes AM, Qi C, and Dullemond CP (2010) *Astrophysical Journal* 723: 1241.
Arras P and Socrates A (2010) *Astrophysical Journal* 714: 1.
Bakos GÁ, Hartman JD, Torres G, et al. (2012) *Astronomical Journal* 144: 19.
Baraffe I, Chabrier G, and Barman T (2010) *Reports on Progress in Physics* 73: 016901.
Baraffe I, Chabrier G, Fortney J, and Sotin C (2014) In: Beuther H, Klessen RS, Dullemond CP, and Henning T (eds.) *Protostars and Planets VI*. University of Arizona Press.
Barban C, Goupil MJ, Rauer H, and Plato 2. 0 Team (2013) In: Jain K, Tripathy SC, Hill F, Leibacher JW, and Pevtsov AA (eds.) *Astronomical Society of the Pacific Conference Series*, vol. 478, p. 115.
Barclay T, Rowe JF, Lissauer JJ, et al. (2013) *Nature* 494: 452.
Batalha NM, Borucki WJ, Bryson ST, et al. (2011) *Astrophysical Journal* 729: 27.
Batalha NM, Rowe JF, Bryson ST, et al. (2013) *The Astrophysical Journal Supplement Series* 204: 24.
Batygin K, Bodenheimer P, and Laughlin G (2009) *Astrophysical Journal* 704: L49.
Batygin K and Stevenson DJ (2010) *Astrophysical Journal* 714: L238.
Batygin K, Stevenson DJ, and Bodenheimer PH (2011) *Astrophysical Journal* 738: 1.
Bean JL, Désert J-M, Kabath P, et al. (2011) *Astrophysical Journal* 743: 92.
Bean JL, Miller-Ricci Kempton E, and Homeier D (2010) *Nature* 468: 669.
Becker JC and Batygin K (2013) *Astrophysical Journal* 778: 100.
Bennett DP, Rhie SH, Nikolaev S, et al. (2010) *Astrophysical Journal* 713: 837.
Berdyugina SV, Berdyugin AV, Fluri DM, and Piirola V (2008) *Astrophysical Journal* 673: L83.
Berta ZK, Charbonneau D, Désert J-M, et al. (2012) *Astrophysical Journal* 747: 35.
Beuzit et al. (2008) SPHERE: a 'Planet Finder' instrument for the VLT. *Proceedings of SPIE* 7014: 701418.
Black DC and Scargle JD (1982) *Astrophysical Journal* 263: 854.
Bodenheimer P, Hubickyj O, and Lissauer JJ (2000) *Icarus* 143: 2.
Bodenheimer P, Laughlin G, and Lin DNC (2003) *Astrophysical Journal* 592: 555.
Bodenheimer P, Lin DNC, and Mardling RA (2001) *Astrophysical Journal* 548: 466.
Boley AC (2009) *Astrophysical Journal* 695: L53.
Bonfils X, Gillon M, Udry S, et al. (2012) *Astronomy and Astrophysics* 546: A27.
Borucki WJ, Koch D, Jenkins J, et al. (2009) *Science* 325: 709.
Brogi M, Snellen IAG, de Kok RJ, et al. (2013) *Astrophysical Journal* 767: 27.
Burrows A and Orton G (2011) *Giant Planet Atmospheres*, pp. 419–440. Tucson, AZ: University of Arizona Press.
Butler RP, Wright JT, Marcy GW, et al. (2006) *Astrophysical Journal* 646: 505.
Carson J, Thalmann C, Janson M, et al. (2013) *Astrophysical Journal* 763: L32.
Casertano S, Lattanzi MG, Sozzetti A, et al. (2008) *Astronomy and Astrophysics* 482: 699.
Chabrier G and Baraffe I (2000) *Annual Review of Astronomy and Astrophysics* 38: 337.
Chambers JE (2001) *Icarus* 152: 205.
Charbonneau D, Allen LE, Megeath ST, et al. (2005) *Astrophysical Journal* 626: 523.
Charbonneau D, Berta ZK, Irwin J, et al. (2009) *Nature* 462: 891.
Charbonneau D, Brown TM, Latham DW, and Mayor M (2000) *Astrophysical Journal* 529: L45.
Charbonneau D, Brown TM, Noyes RW, and Gilliland RL (2002) *Astrophysical Journal* 568: 377.
Chiang EI and Goldreich P (1997) *Astrophysical Journal* 490: 368.
Chiang E and Laughlin G (2013) *Monthly Notices of the Royal Astronomical Society* 431: 3444.
Chiang E and Youdin A (2010) *The Annual Review of Earth and Planetary Sciences:* 38.
Cochran WD, Fabrycky DC, Torres G, et al. (2011) *The Astrophysical Journal Supplement Series* 197: 7.
Cowan NB and Agol E (2008) *Astrophysical Journal* 678: L129.
Cowan NB, Machalek P, Croll B, et al. (2012) *Astrophysical Journal* 747: 82.
Croll B, Albert L, Jayawardhana R, et al. (2011a) *Astrophysical Journal* 736: 78.
Croll B, Lafreniere D, Albert L, et al. (2011b) *Astronomical Journal* 141: 30.
Crossfield IJM, Barman T, Hansen BMS, and Howard AW (2013) *Astronomy and Astrophysics* 559: A33.
Cumming A, Butler RP, Marcy GW, et al. (2008) *Publications of the Astronomical Society of the Pacific* 120: 531.
D'Angelo G, Durisen RH, and Lissauer JJ (eds.) (2011) *Giant Planet Formation*, pp. 319–346. Tucson, AZ: University of Arizona Press.
D'Angelo G, Henning T, and Kley W (2002) *Astronomy and Astrophysics* 385: 647.
D'Angelo G, Kley W, and Henning T (2003) *Astrophysical Journal* 586: 540.
Dawson RI and Fabrycky DC (2010) *Astrophysical Journal* 722: 937.
de Mooij EJW, Brogi M, de Kok RJ, et al. (2012) *Astronomy and Astrophysics* 538: A46.
Deming D, Seager S, Richardson LJ, and Harrington J (2005) *Nature* 434: 740.
Deming D, Seager S, Winn J, et al. (2009) *Publications of the Astronomical Society of the Pacific* 121: 952.
Désert J-M, Bean J, Miller-Ricci Kempton E, et al. (2011) *Astrophysical Journal* 731: L40.
Doyle LR, Carter JA, Fabrycky DC, et al. (2011) *Science* 333: 1602.
Dumusque X, Lovis C, Ségransan D, et al. (2011) *Astronomy and Astrophysics* 535: A55.
Dumusque X, Pepe F, Lovis C, et al. (2012) *Nature* 491: 207.
Durisen RH, Boss AP, Mayer L, et al. (2007) *Protostars and Planets V*, p. 607. Tucson: University of Arizona Press.
Eggleton PP and Kiseleva-Eggleton L (2001) *Astrophysical Journal* 562: 1012.
Enoch B, Cameron AC, Anderson DR, et al. (2011) *Monthly Notices of the Royal Astronomical Society* 410: 1631.
Ercolano B and Koepferl C (2012) *ArXiv e-prints*.
Evans TM, Pont F, Sing DK, et al. (2013) *Astrophysical Journal* 772: L16.
Fabrycky DC, Johnson ET, and Goodman J (2007) *Astrophysical Journal* 665: 754.
Fabrycky DC, Lissauer JJ, Ragozzine D, et al. (2014) *Astrophysical Journal* 146.
Fabrycky DC and Murray-Clay RA (2010) *Astrophysical Journal* 710: 1408.
Fabrycky D and Tremaine S (2007) *Astrophysical Journal* 669: 1298.
Fischer DA and Valenti J (2005) *Astrophysical Journal* 622: 1102.
Ford EB (2006) *Astrophysical Journal* 642: 505.
Fortney JJ, Lodders K, Marley MS, and Freedman RS (2008) *Astrophysical Journal* 678: 1419.
Fortney JJ, Marley MS, and Barnes JW (2007) *Astrophysical Journal* 659: 1661.
Galicher R, Marois C, Zuckerman B, and Macintosh B (2013) *Astrophysical Journal* 769: 42.
Gammie CF (2001) *Astrophysical Journal* 553: 174.
Gaudi BS (2012) *Annual Review of Astronomy and Astrophysics* 50: 411.
Gold T and Soter S (1969) *Icarus* 11: 356.
Goldreich P and Ward WR (1973) *Astrophysical Journal* 183: 1051.
Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) *Nature* 435: 466.
Gonzalez G (1997) *Monthly Notices of the Royal Astronomical Society* 285: 403.
Gonzalez G (1999) *Monthly Notices of the Royal Astronomical Society* 308: 447.
Greenberg R, Hartmann WK, Chapman CR, and Wacker JF (1978) *Icarus* 35: 1.
Greenzweig Y and Lissauer JJ (1992) *Icarus* 100: 440.
Grillmair CJ, Burrows A, Charbonneau D, et al. (2008) *Nature* 456: 767.
Gu P-G and Ogilvie GI (2009) *Monthly Notices of the Royal Astronomical Society* 395: 422.
Guillot T and Showman AP (2002) *Astronomy and Astrophysics* 385: 156.
Haisch KE Jr., Lada EA, and Lada CJ (2001) *Astrophysical Journal* 553: L153.

Hansen BMS and Murray N (2012) *Astrophysical Journal* 751: 158.
Hansen CJ, Schwartz JC, and Cowan NB (2014) *MNRAS* 444: 3632.
Hayashi C (1981) *Progress of Theoretical Physics Supplement* 70: 35–53.
Helled R, Bodenheimer P, Podolak M, et al. (2013) In: Beuther H, Klessen RS, Dullemond CP, and Henning T (eds.) *Protostars and Planets VI*. University of Arizona Press.
Henry GW, Marcy GW, Butler RP, and Vogt SS (2000) *Astrophysical Journal* 529: L41.
Howard AW, Johnson JA, Marcy GW, et al. (2011) *Astrophysical Journal* 730: 10.
Howard AW, Sanchis-Ojeda R, Marcy GW, et al. (2013) A rocky composition for an Earth-sized exoplanet. *Nature* 503(7476): 381–384.
Hubbard WB, Podolak M, and Stevenson DJ (1995) In: Cruikshank DP, Matthews MS, and Schumann AM (eds.) *Neptune and Triton*, 109–138.
Hubickyj O, Bodenheimer P, and Lissauer JJ (2005) *Icarus* 179: 415.
Inaba S, Wetherill GW, and Ikoma M (2003) *Icarus* 166: 46.
Jacob WS (1855) *Monthly Notices of the Royal Astronomical Society* 15: 228.
Johansen A, Youdin AN, and Lithwick Y (2012) *Astronomy and Astrophysics* 537: A125.
Jontof-Hutter D, Lissauer JJ, Rowe JF, and Fabrycky D (2013) In: *AAS/Division for Planetary Sciences Meeting Abstracts*, Vol. 45, 300.07.
Jontof-Hutter D, Lissauer JJ, Rowe JF, and Fabrycky DC (2014) *Astrophysical Journal* 785: 15.
Jurić M and Tremaine S (2008) *Astrophysical Journal* 686: 603.
Kains N, Street RA, Choi J-Y, et al. (2013) *Astronomy and Astrophysics* 552: A70.
Kalas P, Graham JR, Chiang E, et al. (2008) *Science* 322: 1345.
Kenyon SJ, Currie T, and Bromley BC (2014) *ApJ* 786: 70.
Kley W and Nelson RP (2012) *Annual Review of Astronomy and Astrophysics* 50: 211.
Knutson HA, Charbonneau D, Allen LE, et al. (2007) *Nature* 447: 183.
Knutson HA, Charbonneau D, Cowan NB, et al. (2009a) *Astrophysical Journal* 703: 769.
Knutson HA, Charbonneau D, Cowan NB, et al. (2009b) *Astrophysical Journal* 690: 822.
Knutson HA, Howard AW, and Isaacson H (2010) *Astrophysical Journal* 720: 1569.
Knutson HA, Lewis N, Fortney JJ, et al. (2012) *Astrophysical Journal* 754: 22.
Kokubo E and Ida S (1998) *Icarus* 131: 171.
Konacki M and Wolszczan A (2003) *Astrophysical Journal* 591: L147.
Konopacky QM, Barman TS, Macintosh BA, and Marois C (2013) *Science* 339: 1398.
Kramm U, Nettelmann N, Fortney JJ, Neuhäuser R, and Redmer R (2012) *Astronomy and Astrophysics* 538: A146.
Kreidberg L, Bean J, Désert J, et al. (2013) In: *American Astronomical Society Meeting Abstracts*, Vol. 221, 224.03.
Krist JE, Stapelfeldt KR, Bryden G, and Plavchan P (2012) *Astronomical Journal* 144: 45.
Kuzuhara M, Tamura M, Kudo T, et al. (2013) *Astrophysical Journal* 774: 11.
Lagrange A-M, Bonnefoy M, Chauvin G, et al. (2010) *Science* 329: 57.
Langton et al. (2014) *Astrophysical Journal*, Submitted.
Laskar J (2000) *Physical Review Letters* 84: 3240.
Latham DW, Stefanik RP, Mazeh T, Mayor M, and Burki G (1989) *Nature* 339: 38.
Laughlin G, Crismani M, and Adams FC (2011) *Astrophysical Journal* 729: L7.
Lazio TJW, Shankland PD, Farrell WM, and Blank DL (2010) *Astronomical Journal* 140: 1929.
Léger A, Rouan D, Schneider J, et al. (2009) *Astronomy and Astrophysics* 506: 287.
Levrard B, Correia ACM, Chabrier G, et al. (2007) *Astronomy and Astrophysics* 462: L5.
Lewis NK, Knutson HA, Showman AP, et al. (2013) *Astrophysical Journal* 766: 95.
Lin DNC, Bodenheimer P, and Richardson DC (1996) *Nature* 380: 606.
Lin DNC and Papaloizou JCB (1979) *Monthly Notices of the Royal Astronomical Society* 186: 799.
Lin DNC and Papaloizou J (1986) *Astrophysical Journal* 309: 846.
Line MR, Zhang X, Vasisht G, et al. (2012) *Astrophysical Journal* 749: 93.
Lissauer JJ (1987) *Icarus* 69: 249.
Lissauer JJ (1995) *Icarus* 114: 217.
Lissauer JJ (2006) Theories of planet formation: Future prospects. In: S Udry, W Benz, and R von Steiger (eds.) Planetary Systems and Planets in Systems, p. 249. Bern, Switzerland: International Space Science Institute. ISBN-13 978-92-9221-935-2.
Lissauer JJ, Fabrycky DC, Ford EB, et al. (2011a) *Nature* 470: 53.
Lissauer JJ, Hubickyj O, D'Angelo G, and Bodenheimer P (2009) *Icarus* 199: 338.
Lissauer JJ, Marcy GW, Bryson ST, et al. (2014a) *Astrophysical Journal* 784: 44.
Lissauer JJ, Dawson RI, and Tremaine S (2014b) Advances in exoplanet science from Kepler. *Nature* 513: 336–344.
Lissauer JJ, Ragozzine D, Fabrycky DC, et al. (2011b) *The Astrophysical Journal Supplement Series* 197: 8.
Lissauer JJ and Stevenson DJ (2007) *Protostars and Planets V*, p. 591.
Lubow SH and Ida S (2011) *Planet Migration*, pp. 347–371. Tucson, AZ: University of Arizona Press.
Macintosh B, Graham JR, Ingraham P, et al. (2014) *PNAS* 11112661.
Madhusudhan N (2012) *Astrophysical Journal* 758: 36.
Madhusudhan N, Knutson H, Fortney J, and Barman T (2014) In: Beuther H, Klessen RS, Dullemond CP, and Henning T (eds.) *Protostars and Planets VI*. University of Arizona Press.
Madhusudhan N and Seager S (2010) *Astrophysical Journal* 725: 261.
Mamajek EE (2009) In: Usuda T, Tamura M, and Ishii M (eds.) *American Institute of Physics Conference Series American Institute of Physics Conference Series*, vol. 1158, pp. 3–10.
Marcy GW and 102 colleagues (2014) Masses, radii, and orbits of small kepler planets: the transition from gaseous to rocky planets. *Astrophys. J. Suppl.* 210: 20.
Marcy GW, Isaacson H, Howard AW, et al. (2014) *The Astrophysical Journal Supplement Series* 210: 20.
Mardling RA (2007) *Monthly Notices of the Royal Astronomical Society* 382: 1768.
Mardling RA and Lin DNC (2002) *Astrophysical Journal* 573: 829.
Marley MS, Fortney JJ, Hubickyj O, Bodenheimer P, and Lissauer JJ (2007) *Astrophysical Journal* 655: 541.
Marois C, Macintosh B, Barman T, et al. (2008) *Science* 322: 1348.
Marois C, Zuckerman B, Konopacky QM, Macintosh B, and Barman T (2010) *Nature* 468: 1080.
Maxted PFL, Anderson DR, Doyle AP, et al. (2013) *Monthly Notices of the Royal Astronomical Society* 428: 2645.
Mayor M, Lovis C, Pepe F, Ségransan D, and Udry S (2011a) *Astronomische Nachrichten* 332: 429.
Mayor M, Marmier M, Lovis C, et al. (2011b) *ArXiv e-prints*.
Mayor M, Udry S, Lovis C, et al. (2009) *Astronomy and Astrophysics* 493: 639.
McLaughlin DB (1924) *Astrophysical Journal* 60: 22.
Megeath ST, Allen LE, Gutermuth RA, et al. (2004) *The Astrophysical Journal Supplement Series* 154: 367.
Meschiari S, Wolf AS, Rivera E, et al. (2009) *Publications of the Astronomical Society of the Pacific* 121: 1016.
Meyer MR, Hillenbrand LA, Backman DE, et al. (2004) *The Astrophysical Journal Supplement Series* 154: 422.
Miller N and Fortney JJ (2011) *Astrophysical Journal* 736: L29.
Mizuno H (1980) *Progress of Theoretical Physics* 64: 544.
Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) *Nature* 435: 462.
Morton TD and Johnson JA (2011) *Astrophysical Journal* 738: 170.
Movshovitz N, Bodenheimer P, Podolak M, and Lissauer JJ (2010) *Icarus* 209: 616.
O'Brien DP, Morbidelli A, and Levison HF (2006) *Icarus* 184: 39.
Ogihara M, Ida S, and Morbidelli A (2007) *Icarus* 188: 522.
Peale SJ (1997) *Icarus* 127: 269.
Pepe F, Cameron AC, Latham DW, et al. (2013) An Earth-sized planet with an Earth-like density. *Nature* 503(7476): 377–380.
Perryman M (2011) In: Seager S (ed.) *The Exoplanet Handbook*. Tucson AZ: University of Arizona Press.
Pollack JB, Hubickyj O, Bodenheimer P, et al. (1996) *Icarus* 124: 62.
Rameau J, Chauvin G, Lagrange A-M, et al. (2013) *Astrophysical Journal* 779: L26.
Rasio FA and Ford EB (1996) *Science* 274: 954.
Rauscher E and Menou K (2013) *Astrophysical Journal* 764: 103.
Ricker, G. R., Latham, D. W., Vanderspek, R. K., et al. (2010) In: Bulletin of the American Astronomical Society, Vol. 42, American Astronomical Society, Meeting Abstracts #215, 450.06.
Rivera EJ, Lissauer JJ, Butler RP, et al. (2005) *Astrophysical Journal* 634: 625.
Rogers LA, Bodenheimer P, Lissauer JJ, and Seager S (2011) *Astrophysical Journal* 738: 59.
Rogers L and Seager S (2010) A framework for quantifying the degeneracies of exoplanet interior compositions. *Astrophysical Journal* 712: 974.
Rossiter RA (1924) *Astrophysical Journal* 60: 15.
Rowe JF and 28 colleagues (2014) Validation of kepler's multiple planet candidates. III. light curve analysis and announcement of hundreds of new multi-planet systems. *The Astrophys. J.* 784: 45.
Rowe JF, Matthews JM, Seager S, et al. (2008) *Astrophysical Journal* 689: 1345.
Safronov VS (1969) *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*. Jerusalem: IPST.
Sanchis-Ojeda R, Rappaport S, Winn JN, Levine A, Kotson MC, Latham DW, and Buchhave LA (2013) Transits and occultations of an earth-sized planet in an 8.5 hr orbit. *Astrophys. J.* 774: 54.
Santerne A, Fressin F, Díaz RF, et al. (2013) *Astronomy and Astrophysics* 557: A139.
Santos NC, Israelian G, and Mayor M (2001) *Astronomy and Astrophysics* 373: 1019.
Seager S and Deming D (2010) *Annual Review of Astronomy and Astrophysics* 48: 631.
Seager S, Whitney BA, and Sasselov DD (2000) *Astrophysical Journal* 540: 504.
See TJJ (1896) *Astronomical Journal* 16: 17.
Showman AP, Fortney JJ, Lewis NK, and Shabram M (2013) *Astrophysical Journal* 762: 24.

Showman AP, Fortney JJ, Lian Y, et al. (2009) *Astrophysical Journal* 699: 564.
Showman AP and Polvani LM (2011) *Astrophysical Journal* 738: 71.
Shu FH (1992) The physics of astrophysics. *Gas Dynamics*, vol. II. Mill Valley, CA: University Science Books.
Sigurdsson S, Richer HB, Hansen BM, Stairs IH, and Thorsett SE (2003) *Science* 301: 193.
Smith AMS, Anderson DR, Bouchy F, et al. (2013) *Astronomy and Astrophysics* 552: A120.
Smith BA and Terrile RJ (1984) *Science* 226: 1421.
Snellen IAG, de Kok RJ, de Mooij EJW, and Albrecht S (2010) *Nature* 465: 1049.
Sotin C, Grasset O, and Mocquet A (2007) *Icarus* 191: 337.
Sotin C, Jackson JM, and Seager S (2011) Terrestrial planet interiors. In: Seager S (ed.) Exoplanets, pp. 375–395. Tucson, AZ: University of Arizona Press. ISBN: 978-0-8165-2945-2.
Sousa SG, Santos NC, Mayor M, et al. (2008) *Astronomy and Astrophysics* 487: 373.
Southworth J (2010) *Monthly Notices of the Royal Astronomical Society* 408: 1689.
Steffen JH, Ragozzine D, Fabrycky DC, et al. (2012) *Proceedings of the National Academy of Science* 109: 7982.
Sterne TE (1939) *Monthly Notices of the Royal Astronomical Society* 99: 451.
Strom SE, Edwards S, and Skrutskie MF (1993) In: Levy EH and Lunine JI (eds.) *Protostars and Planets III*, pp. 837–866.
Sumi T, Kamiya K, Bennett DP, et al. (2011) *Nature* 473: 349.
Swain MR, Tinetti G, Vasisht G, et al. (2009a) *Astrophysical Journal* 704: 1616.
Swain MR, Vasisht G, Tinetti G, et al. (2009b) *Astrophysical Journal* 690: L114.
Terebey S, Shu FH, and Cassen P (1984) *Astrophysical Journal* 286: 529.
Tinetti G, Encrenaz T, and Coustenis A (2013) *The Astronomy and Astrophysics Review* 21: 63.
Torres G, Fressin F, Batalha NM, et al. (2011) *Astrophysical Journal* 727: 24.
Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) *Nature* 435: 459.
van de Kamp P (1963) *Astronomical Journal* 68: 515.
van de Kamp P (1969) *Astronomical Journal* 74: 757.
Vogt SS, Wittenmyer RA, Butler RP, et al. (2010) *Astrophysical Journal* 708: 1366.
Walsh KJ, Morbidelli A, Raymond SN, O'Brien DP, and Mandell AM (2011) *Nature* 475: 206.
Ward WR (1997) *Icarus* 126: 261.
Weidenschilling SJ (1977) *Monthly Notices of the Royal Astronomical Society* 180: 57.
Weidenschilling SJ and Cuzzi JN (1993) In: Levy EH and Lunine JI (eds.) *Protostars and Planets III*, 1031–1060.
Weiss LM, Marcy GW, Rowe JF, et al. (2013) *Astrophysical Journal* 768: 14.
Werner MW, Roellig TL, Low FJ, et al. (2004) *The Astrophysical Journal Supplement Series* 154: 1.
Wetherill GW and Stewart GR (1989) *Icarus* 77: 330.
Wiktorowicz SJ (2009) *Astrophysical Journal* 696: 1116.
Williams JP and Cieza LA (2011) *Annual Review of Astronomy and Astrophysics* 49: 67.
Winn JN, Noyes RW, Holman MJ, et al. (2005) *Astrophysical Journal* 631: 1215.
Wolszczan A and Frail DA (1992) *Nature* 355: 145.
Woolfson MM (1993) *Quarterly Journal of the Royal Astronomical Society* 34: 1.
Wright JT, Marcy GW, Howard AW, et al. (2012) *Astrophysical Journal* 753: 160.
Wu Y and Goldreich P (2002) *Astrophysical Journal* 564: 1024.
Wu Y and Lithwick Y (2012) *ApJ* 756: 11.
Wu Y and Murray N (2003) *Astrophysical Journal* 589: 605.
Wuchterl G, Guillot T, and Lissauer JJ (2000) In: *Protostars and Planets IV*, p. 1081.
Wyatt MC (2008) *Annual Review of Astronomy and Astrophysics* 46: 339.
Young RE (2003) *The New Educational Review* 47: 1.

10.22 Mission Analysis Issues for Planetary Exploration Missions

Y Langevin, CNRS/Université Paris Sud, Orsay, France

© 2015 Elsevier B.V. All rights reserved.

10.22.1	**Scientific Rationale and Present Status of Planetary Exploration Missions**	695
10.22.2	**Energy Requirements and Mass Budgets for Planetary Missions**	697
10.22.2.1	Leaving the Gravitational Influence of the Earth	697
10.22.2.2	Flyby Missions	699
10.22.2.3	Rendezvous Missions	700
10.22.2.3.1	Rendezvous missions to natural satellites of giant planets	700
10.22.2.3.2	Rendezvous missions to comets and Earth-crossing asteroids	700
10.22.3	**Remote-Sensing and In Situ Missions to Venus and Mars**	701
10.22.3.1	Orbiter Missions Not Implementing Aerocapture	702
10.22.3.2	In Situ Missions and Orbiter Missions Implementing Aerocapture	703
10.22.4	**Orbit Evolution Around Planetary Bodies**	704
10.22.5	**In Situ Missions to Atmosphereless Bodies**	704
10.22.6	**Gravity-Assist Missions: Giant Planets, Mercury, Asteroid, and Comet Rendezvous**	706
10.22.7	**Advanced Propulsion Systems: Solar Sail and Ion Propulsions**	711
10.22.7.1	Solar Sail Propulsion	712
10.22.7.2	Ion Propulsion	712
10.22.8	**The Specific Challenges of Sample Return Missions**	714
References		716

10.22.1 Scientific Rationale and Present Status of Planetary Exploration Missions

For centuries, major advances in planetary science had to rely on ground-based observations. The increasing size of telescopes up to the present 8 m class observatories (e.g., VLT and Keck) and the increasing sophistication of focal plane instruments have made major advances possible in terms of wavelength coverage and spectral and spatial resolution for observing solar system bodies. Two developments have much enhanced the potential of ground-based observations over the last 20 years: the development of adaptive optics (Beckers, 1993), which brings telescopes close to their diffraction limits by correcting for atmospheric turbulence, and interferometry, which increases the effective aperture, improving spatial resolution. New results can therefore be expected from the next generation of ground-based observatories in the 2010–20 time frame, in particular ALMA (in the sub-mm wavelength range, first light in 2012), extremely large telescopes (aperture 30–40 m, scheduled for 2020), and the Square Kilometer Array (at radio wavelengths). Space observatories are not limited to the transmission windows of the Earth atmosphere. The most prominent examples for planetary science have been the Hubble Space Telescope in the visible and near-IR range, ISO and SIRTF in the IR range, Herschel (launched in 2009) in the sub-mm wavelength range, and FUSE in the UV range. The next major step will be the launch of JWST, the successor to Hubble with a wavelength range extending to the middle IR, in 2018.

An important perspective for planetary science that goes beyond our solar system is the detection and characterization of other planetary systems, which cannot yet be considered for in situ investigations (**Chapter 10.21**). Two space observatories dedicated to exoplanetary transits, COROT (CNES–ESA, launched in 2006) and Kepler (NASA, launched in 2009), have been highly successful in discovering exoplanets down to a few Earth masses. GAIA (ESA), successfully launched in December 2013, should identify thousands of giant planets through the displacement of the central star. The next steps are already being prepared, with three transit missions selected in 2013 and 2014. A small ESA mission, Cheops, approved in October 2013, will accurately determine the radius of exoplanets orbiting bright stars. Plato (ESA) will combine long-duration transit observations with ground-based radial velocity determinations so as to determine the radius and mass of exoplanets, hence their density. TESS is a NASA mission dedicated to finding targets for transit spectroscopy, the next step in exoplanet characterization, which is one of the goals of JWST (launch now scheduled in 2018).

While recognizing the importance of these contributions, it is however clear that most breakthroughs in solar system science have resulted and will continue in the future to result from space missions to other solar system bodies, starting from the landmark Luna 3 mission to the Moon, which revealed for the first time its hemisphere facing away from the Earth. Similarly to space observatories, remote-sensing planetary missions are not limited to the transmission windows of the atmosphere of the Earth. They can achieve much higher spatial resolutions even with relatively small optical devices (a typical planetary mission payload ranges from 50 to 150 kg in mass) as they get very close to the target body. Let us consider a 10 m class telescope operating at 1 μm. Its maximum resolution is 10^{-7} rad assuming that the diffraction limit is reached using adaptive optics. The best observing conditions are at quadrature for Mercury and Venus (maximum angular distance from the Sun) and at opposition for outer bodies. Distances therefore range from 55 million km (Mars at the best oppositions) to

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00180-9

several hundred million km. With a space mission, one can easily implement a camera with an aperture of at least 10 cm (a 70 cm aperture camera, HiRISE, is orbiting Mars since November 2006), corresponding to a diffraction-limited resolution of 10^{-5} rad or better. The closest distance for a flyby or an orbiter is typically 300 km for most planets and asteroids, 30 000 km for Jupiter, and 80 000 km for Saturn (outer edge of the A ring). For Mars, the resolution of a camera with a 10–20 kg mass is in the meters range instead of at best 5.5 km from the Earth. Therefore, flyby missions (over a short time interval) and orbiter missions provide improvements by several orders of magnitudes over the best ground-based or Earth-orbiting facilities. Orbiter missions make it possible to monitor temporal variations at high spatial resolution. Such missions also provide unique opportunities for characterizing the environment of planetary bodies in terms of particles, fields, and their interactions with the solar wind. By implementing in situ missions (landers, rovers, and airborne modules if there is an atmosphere), it is possible to bring to bear the full range of techniques used in Earth sciences: chemical and mineralogical analysis, seismology (which is best implemented with a network of stations), atmospheric sensors, etc. A comprehensive review of experimentation for planetary missions is provided in the next chapter (**Chapter 10.23**). Eventually, sample return missions make it possible to implement powerful analysis techniques, which are still required today for very high-precision measurements of trace elements and isotopic compositions (radiochronology and isotopic anomalies). This may change in the future considering the remarkable improvements of in situ analysis techniques over the last 20 years (e.g., mass spectrometry with a resolution of 1000 or more, implemented on the Rosetta comet rendezvous mission of ESA only 20 years after it became available in the laboratory).

As discussed in, for example, Langevin (2001), the sequence of missions to a given planetary body in terms of increasing complexity is therefore the following:

- First, exploration by a flyby mission
- Comprehensive characterization by remote-sensing techniques on orbiter missions
- In situ investigations (landers, rovers, descent probes, balloons, airplanes, etc.)
- Sample return missions

Possibly even more so than the 1st edition in 2006, the 2nd edition of this book corresponds to a timely period for assessing the status of planetary exploration and outlining the next perspectives. All classes of solar system bodies with the exception of Kuiper Belt objects (KBOs, beyond Neptune) have been visited at least once by flyby missions. The last gap will be filled by the New Horizons NASA mission, which will encounter Pluto, one of the largest KBOs, in 2015. Similarly, the remote-sensing phase is well under way, with the highly successful missions to Mars (MGS, Mars Odyssey, and Mars Reconnaissance Orbiter (MRO) for NASA and Mars Express for ESA), the Cassini–Huygens mission to the Saturn–Titan system (NASA–ESA), the Venus Express mission to Venus (ESA), and the Messenger mission (NASA, launched in 2005, orbiting Mercury since March 2011) and the more ambitious BepiColombo mission to Mercury (ESA/JAXA, launch scheduled in 2016). Five lunar orbiter missions launched from 2007 to 2010 (Kaguya, Japan; Lunar Reconnaissance Orbiter, NASA; Chandrayaan, India; and Chang'e 1 and 2, China) provided a comprehensive remote-sensing coverage of the Moon. The outstanding results of this extensive set of ambitious planetary exploration missions are discussed in other sections of Volume 10, in particular **Chapters 10.02** and **10.05** for the terrestrial planets, **Chapter 10.16** for the giant planets, and **Chapter 10.18** for icy satellites. The JUICE mission selected by ESA in 2012 for a launch in 2022 will perform a comprehensive remote-sensing investigation of Jupiter, its environment, and its satellites, in particular Ganymede and Callisto. For small bodies, one can mention the rendezvous missions toward comet Churyumov–Gerasimenko (Rosetta, ESA, rendezvous in mid-2014), near-Earth asteroids Eros (NEAR, NASA) and Itokawa, and a primitive NEA (Hayabusa and Hayabusa 2, Japan). The Dawn mission (NASA), launched in 2007, has successfully orbited asteroid 4 Vesta for more than 1 year (July 2011–September 2012), and it is now on its way to Ceres, the largest main belt asteroid, which will be reached in 2015. With this impressive set of orbital remote-sensing missions (in operation or approved), comprehensive information has been or will be obtained on six of the eight planets (Uranus and Neptune being left out), four of the seven planet-sized satellites (the Moon, Ganymede, Callisto, and Titan), one cometary nucleus, the two largest main belt asteroids, and three near-Earth asteroids.

In situ investigations suffered some setbacks in the first years of the twenty-first century (failure of the Beagle 2 Mars lander and descoping of a Mercury lander on BepiColombo). However, remarkable successes have been obtained by the Huygens descent probe in the atmosphere of Titan (ESA), which provided information all the way down to the surface in January 2005, and the Mars Exploration Rovers of NASA, 'Spirit' and 'Opportunity,' which started their investigations in early 2004 and much exceeded their 3 months' lifetime (Opportunity was still operating in late 2012). A very large Mars rover ('Curiosity,' 800 kg) was launched in November 2011 by NASA, and it has been investigating the interior of the Gale crater since its successful deployment in August 2012. ExoMars, the rover mission to Mars in the framework of the exploration program of ESA, initially scheduled in 2013, is now considered for a 2018 launch at the earliest. The Rosetta lander, Philae, has landed in November 2014 as the first lander ever on a cometary nucleus. China, Japan, and India are presently studying lunar lander and rover missions for the 2014–20 time frame.

For 30 years, lunar dust and rocks from Apollo and Luna had remained the only samples brought back from another planetary body. Samples of cometary dust have been successfully recovered in early 2006 from comet Wild-1 (Stardust, NASA), with intriguing results on early mixing in the protosolar nebula. The Japanese Hayabusa mission has succeeded in bringing back dust grains from asteroid Itokawa, and a reflight to a primitive asteroid has been launched in December 2014. Mars sample return missions are now considered beyond the 2020 time frame.

Given the remarkable success and breadth of scope of presently flying and approved remote-sensing missions, it is likely that planetary missions beyond the 2020 time frame will focus on in situ investigations and sample returns. Orbiter missions will, however, continue to be useful as relays for in situ modules, specific science objectives, and extended time coverage. These considerations are important for the purpose of this chapter, as the resources required are increasingly larger for

each step in planetary exploration, even if the presence of an atmosphere around the target body can in part alleviate for these higher energy requirements. The following sections present the major constraints on mission design for planetary exploration. The basic issues of astrodynamics are presented in a reference textbook (Bate et al., 1971). More recent textbooks have been published by Battin (1999) and Vallado (2001). A discussion of these questions focused on mission design techniques is also available (Langevin, 2005). A review of the most recent advances in mission analysis is provided by Marsden and Ross (2006).

10.22.2 Energy Requirements and Mass Budgets for Planetary Missions

10.22.2.1 Leaving the Gravitational Influence of the Earth

Apart from missions to the Moon, all planetary missions have to leave the gravitational influence of the Earth. So as to compare the different types of missions, one can consider as a starting point a near-circular orbit around the Earth at an altitude of ~330 km, which lies beyond the upper reaches of the Earth atmosphere. A large velocity increment can then be provided in as short a time as possible by an upper stage, so as to exceed the escape velocity. This strategy was used for both Mars Express (June 2003) and Venus Express (November 2005) with the Soyuz launcher and its Fregat upper stage.

At this altitude, the velocity on a circular orbit (circular velocity, V_c) 7.71 km s^{-1} and the escape velocity V_{esc}, which is always $\sqrt{2}$ times larger than V_c, are 10.90 km s^{-1}. Velocity increments lower than 3.19 km s^{-1} provided by the upper stage will therefore result in bound orbits, and velocity increments larger than 3.19 km s^{-1} will result in interplanetary trajectories. The interplanetary departure velocity V_{inf} can easily be derived from energy conservation arguments:

$$V_{inf}^2 = V_{per}^2 - V_{esc}^2 \quad [1]$$

where V_{per} is the velocity at pericenter after the upper-stage burn.

The resources required for each class of trajectory can then be compared from the velocity increment ('ΔV'), which is required in addition to that that would be needed to reach a bound orbit with a semimajor axis of 320 000 km, which goes much beyond that of the Moon (Table 1).

It is important to note that the ΔV required from the upper stage to reach the Moon (3.14 km s^{-1}) and to leave the Earth with a 3 km s^{-1} relative velocity (3.61 km s^{-1}) are similar (460 m s^{-1} difference). Very large departure velocities (9 km s^{-1}) require additional velocity increments similar to the 3.19 km s^{-1}, which are already needed to escape from a circular orbit.

These considerations have a major impact on the mass available for interplanetary modules. This directly derives from the basic physical constraints on rockets, first discussed by K. Tsiolkovky in the last years of the nineteenth century. An upper stage, similarly to any rocket, ejects exhaust gas at a large velocity V_{ej} (~3.25 km s^{-1} for the Fregat upper stage and ~4.1 km s^{-1} for the Centaur cryogenic upper stage). From the momentum balance, we can write

$$M(t)\mathrm{d}v = \mathrm{d}M V_{ej} \quad [2]$$

where $M(t)$ is the mass at a given time during the thrust. Integrating eqn [2] over the full duration of the thrust results in the basic constraint of spaceflight, the 'rocket equation' (Tsiolkovsky, 1903):

$$\Delta V = V_{final} - V_{initial} = V_{ej}\, In(M_i/M_f) \quad [3]$$

where M_i and M_f are the initial and final masses of the upper stage with its payload and ΔV is the total velocity increment.

In the case of the Soyuz–Fregat launch strategy used for Mars Express and Venus Express, the equivalent mass, which could be injected in a near-circular orbit, is 7200 kg, while the dry mass of the Fregat stage with the payload attachment device is 1150 kg. It should be noted that this 'useless' dry mass ends up with the same velocity as the spacecraft itself; hence, it has to be subtracted from M_f so as to obtain the 'useful' mass injected into a given orbit.

Figure 1 provides the useful mass as derived from the rocket equation, which can be launched at a given departure velocity

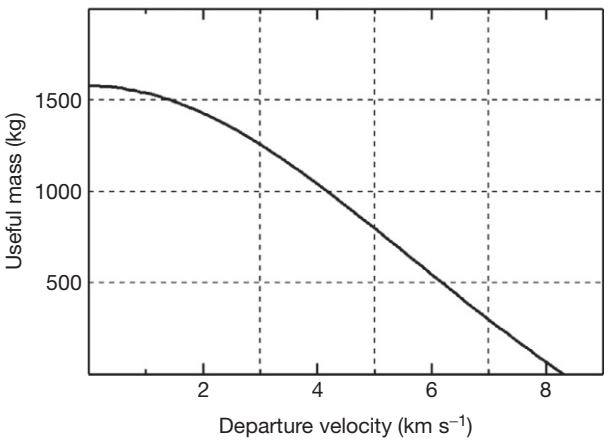


Figure 1 Useful mass as a function of departure velocity with the Soyuz–Fregat combination used for the launch of Mars Express in 2003, not taking into account gravity losses. The departure velocity of Mars Express was ~3 km s^{-1}; hence, the available useful mass was ~1250 kg.

Table 1 ΔV requirements on the upper-stage burn for each class of trajectories

Type of trajectory	V_{per} *(km s^{-1})*	*Upper stage* ΔV *(km s^{-1})*	ΔV *over bound orbit (km s^{-1})*
Moon-crossing orbit	10.85	3.14	0.00
Marginal escape	10.90	3.19	0.05
Departure at 3 km s^{-1}	11.31	3.61	0.46
Departure at 5 km s^{-1}	12.00	4.29	1.15
Departure at 7 km s^{-1}	12.96	5.25	2.11
Departure at 9 km s^{-1}	14.14	6.43	3.29

from the Earth–Moon system with Soyuz–Fregat in its Mars Express configuration. It is clear that there is very little available mass for escape velocities larger than 6 km s^{-1}, while departure velocities lower than 3.5 km s^{-1} result in significant mass budgets even with a relatively small (and cheap!) launcher. This is a slightly optimistic view, as it considers that the Fregat burn is performed at a 330 km altitude. However, even with the powerful thruster of Fregat (20 000 N, providing an acceleration of 2.8 m s^{-2} at full load and 10 m s^{-2} at the end of the thrust arc), the thrust extends over several hundred seconds so as to reach escape velocity. Part of the thrust is therefore provided when the upper stage is at higher altitudes. Close to escape, the energy provided by each m s^{-1} of ΔV is proportional to the local escape velocity. At an altitude of 2000 km, the local escape velocity is 9.75 km s^{-1} instead of 10.90 km s^{-1} at an altitude of 330 km s^{-1}; hence, the efficiency of the thrust in terms of energy has been reduced by 10%. This penalty incurred with extended thrust arcs is called a 'gravity loss.' It is increasingly severe for very large velocity increments, which require longer thrust arcs. As an example, the actual performance of the Soyuz–Fregat drops to zero slightly below 8 km s^{-1}, instead of 8.15 km s^{-1} as in Figure 1. In practical terms, a single stage such as Fregat can provide at best a velocity increment that is ~1.5 times larger than the exhaust velocity. For spacecraft propulsion systems, the practical limit is lower, in the range of the exhaust velocity, as the initial mass is that at departure and some useful mass must remain for the scientific mission.

Two alternative launch strategies with an intermediate bound orbit can be considered:

- For low departure velocities, it is possible to effectively use the Moon, which is a very large satellite when compared to its primary. A Moon gravity assist (MGA, see Section 10.22.6) from a bound orbit can provide a departure velocity close to 1.4 km s^{-1}. If the Moon is in a favorable location, this strategy provides for free the launch capability of the Moon-crossing orbit (1610 kg) instead of 1500 kg as in Figure 1. There are severe limitations to this strategy, as the departure direction must be close to the orbit plane of the Moon. The 7% improvement in the mass budget is therefore only of interest in very specific cases (ion propulsion missions, see Section 10.22.7)
- For large departure velocities, it is not effective to carry along the large dry mass of the Fregat upper stage (1150 kg), which then exceeds the useful mass. It is interesting to consider launching as much mass as possible (1610 kg) into a Moon-crossing elliptical orbit. The spacecraft thruster (or a dedicated propulsion stage) can then be used as a new stage at the next perigee. This indirect launch strategy is discussed by Farquhar and Dunham (1999). The required ΔV at perigee can be derived from the fourth column in Table 1, which indicates the difference between the perigee velocity of the Moon-crossing orbit and that required for departure at a given velocity. Interplanetary spacecraft thrusters provide typically 400 N, much less than the 20 000 N of Fregat. The exhaust velocity is only slightly smaller (3.1 km s^{-1} instead of 3.28 km s^{-1}). For large departure velocities, the thrust arc extends far before and after perigee, and gravity losses begin to offset the advantage of jettisoning the large Fregat upper stage. For Mars Express (departure at 3 km s^{-1}), the spacecraft thruster has to provide 460 m s^{-1} (actually 500 m s^{-1} due to gravity losses). The final mass after the spacecraft thruster burn is then 1370 kg, 130 kg more than with a direct injection with Fregat. One should, however, take into account that 240 kg of additional fuel has to be carried, with a spacecraft mass at separation of 1610 kg. This requires larger tanks and a more robust structure than the 'direct injection' spacecraft with its 1240 kg mass. For each additional kg of fuel, 0.1–0.14 kg of tanks and structure have to be added to the spacecraft (this ratio is called the 'tankage factor'). With 240 kg of additional fuel, the penalty in dry mass is ~30 kg; hence, the benefit of such a launch strategy is reduced to 100 kg, which is still quite significant. For a departure velocity of 6 km s^{-1}, the bound orbit launch strategy increases the available mass after departure from 500 to 700 kg, even taking into account gravity losses. For departure velocities ranging from 3 to 6 km s^{-1}, there is therefore a complex trade-off between mass budget, complexity, cost, and risk: The 'direct injection' spacecraft is lighter, hence cheaper, and a major maneuver 1 month after launch may not be desirable. A bound orbit launch extends the mission duration by 1–2 months, but extending the launch window and coping with launcher errors become very cheap. Finally, it is important to note that there is a practical limit of ~3 km s^{-1} on onboard ΔV, as this corresponds to the ejection velocity; hence, more than 70% of the spacecraft mass would consist of fuel, tanks, and the corresponding structural mass. Therefore, the bound orbit launch strategy cannot be considered for missions requiring large ΔV capabilities after departure.

Soyuz–Fregat has been launched from Kourou starting in 2011, and new versions of Fregat have been developed. A Moon-crossing orbit capability of 2140 kg with Soyuz–Fregat was considered as the initial baseline for the BepiColombo mission to Mercury (launched in 2013), before this mission exceeded the capability of the Soyuz–Fregat launcher due to major mass increases. For missions requiring more mass, one has to use a more capable launcher. Ariane 5 (ECA version), Atlas 5 (421), Delta 4 (M+), or Proton provides a Moon-crossing orbit capability larger than 4500 kg (3500 kg or more for a departure at 3 km s^{-1}). With such launchers, a direct injection is therefore likely to meet the mass requirements of most solar system exploration missions. The departure velocity is, however, still limited to ~7 km s^{-1} if the upper stage is not cryogenic (Ariane 5 and Proton), as the same equations as those for the Fregat upper stage of Soyuz apply. NASA has a major asset for high-energy missions as it was able to develop a cryogenic upper stage (Centaur) as early as 1963. The ejection velocity of a H_2/O_2 engine is larger than 4 km s^{-1}, compared with 3.1 km s^{-1} for bipropellant thrusters such as that of Fregat, and the decrease in useful mass with departure velocity is slower. This made it possible to implement direct missions to Jupiter, which require departure velocities of at least 9 km s^{-1}: Pioneer 10/11, launched by Atlas/Centaur in 1972/1973, and Voyager 1/2, launched by Titan/Centaur in 1977. Such direct missions to Jupiter have a very short cruise time (1.75 years to Jupiter). However, the useful mass is low: The mass of Pioneer 10 was only 270 kg; hence, this strategy

has until now only been used for flyby missions to the outer solar system, the most recent being 'New Horizons' (NASA), launched in 2006 toward Pluto and the Kuiper Belt.

10.22.2.2 Flyby Missions

Flyby missions only require that the spacecraft leaving Earth crosses the path of the target body. The Moon is a specific case, as it orbits the Earth. It can be visited by a spacecraft on a bound geocentric orbit, with a velocity at perigee lower than the escape velocity (see Table 1). Near-Earth asteroids and several comets have orbits that come quite close to that of the Earth. The velocities relative to the Earth, which are required to reach these bodies (albeit at a high flyby velocity), can be so low (<1.5 km s^{-1}) that a simple lunar swingby from a bound orbit is adequate. The mass budget is then similar to that of a lunar flyby mission. The best example of such a cheap flyby mission is the magnetospheric mission ISEE-3 of NASA, retargeted and renamed ICE (International Comet Explorer) after a lunar flyby, which made it possible to encounter comet Giacobini–Zinner in 1991 (Farquhar, 2001).

Multiple flybys of comets and asteroids can cheaply be implemented using gravity-assist strategies (see Section 10.22.6), thereby enhancing the scientific return (Farquhar et al., 2002). This usually requires retargeting the spacecraft by flying close to a planet. When this planet is the Earth, the period of each successive orbit has to be close to a simple fraction of 1 year. As an example, the trajectory of the ESA probe Giotto on its way to the flyby of comet Halley in 1986 had an orbital period close to 0.75 years, which brought the spacecraft back to the Earth after four orbits and 3 years. It was then possible to adjust the Earth swingby so as to visit comet Schwassmann–Wachmann 3 in 1991. Similarly, USSR's VEGA-1 and VEGA-2 missions used a swingby of Venus (with the deployment of balloons in the atmosphere) to reach comet Halley at nearly the same time as Giotto. The most comprehensive mission of this type was expected to be NASA's CONTOUR. Four flybys of comets and asteroids were planned in the nominal mission, three intermediate Earth swingbys retargeting the spacecraft toward the next target. This mission unfortunately failed at launch in 2001.

All other planetary bodies of interest (planets, natural satellites apart from the Moon, main belt asteroids, and most comets) do not come close to the orbit of the Earth. A transfer trajectory is then required, with a significant departure velocity. Apart from near-Earth objects, minor bodies consist of three main groups: comets, main belt asteroids, and KBOs, which orbit beyond Neptune. Pluto is not a planet, but one of the largest representatives of KBOs. Most minor bodies have large eccentricities and large inclinations. Such characteristics much increase the cost of a flyby mission and even more so that of a rendezvous mission. Some large main belt asteroids, such as 4 Vesta, have relatively low eccentricity and inclination, which is also the case for the seven major planets. Visiting a large natural satellite with a flyby mission has very similar requirements to that of a mission to the planet itself.

A first approximation of the relative complexity of solar system exploration missions can be made by assuming that all planetary orbits are circular, with an orbit in the same plane as the orbit of the Earth. The problem in this coplanar, circular approximation becomes very simple: the cheapest transfer between the Earth and such an idealized planet is an ellipse, which is tangent to the two circles (Hohmann transfer). This applies to both a flyby mission and a rendezvous mission. All velocities in the solar system can be scaled to the orbital velocity of the Earth, $V_e = 29.78$ km s^{-1}, if the semimajor axis of the Earth is considered as the unit of distance (astronomical unit or AU). If R_p is the semimajor axis of the planet in these units, we have

$$A = (1 + R_p)/2 \qquad [4]$$

where A is the semimajor axis of the transfer orbit. The cruise time is one-half of the period of the orbit, which is equal to $A^{1.5}$ (in years) if A is expressed in astronomical units. From Kepler's laws, it is then easy to derive the heliocentric velocities at departure (V_1) and arrival (V_2):

$$V_1 = V_e(2 - 2/A)^{0.5} \quad V_2 = V_e\left(2/R_p - 2/A\right)^{0.5} \qquad [5]$$

The orbital velocity V_p of a planet in a circular orbit is $V_e R_p{}^{-0.5}$. The relative velocity at departure (V_{dep}) and arrival (V_{arr}) is therefore

$$V_{dep} = V_1 - V_e \quad V_{arr} = V_2 - V_e R_p{}^{-0.5} \qquad [6]$$

Table 2 presents the required departure and arrival velocities for selected solar system bodies in the circular and coplanar approximation. Mercury and Vesta depart very significantly from this hypothesis in terms of inclination (Table 2, column 2), and the eccentricity of the orbit of Mercury is large (Table 2, column 3). When comparing Table 2 to Table 1 and Figure 1, Venus and Mars stand out as targets, which are relatively easy to reach, with optimum departure velocities in the range of 2.5–3 km s^{-1} (respectively, backward and forward relative to the orbital motion of the Earth at departure). Even a simple flyby to Vesta is already a challenge, as the departure

Table 2 Orbital elements of selected solar system bodies and main characteristics of the corresponding transfer orbits

Target	*Inclination*	*Eccentricity*	*Semimajor axis (AU)*	*Departure (km s^{-1})*	*Arrival (km s^{-1})*	*Cruise (years)*
Mercury	7.01°	0.206	0.387	−7.53	9.61	0.29
Venus	3.39°	0.007	0.723	−2.50	2.71	0.40
Mars	1.85°	0.093	1.524	2.94	−2.65	0.71
Vesta	7.13°	0.090	2.365	5.51	−4.43	1.09
Jupiter	1.31°	0.049	5.203	8.80	−5.64	2.73
Saturn	2.49°	0.056	9.539	9.60	−5.44	6.04

The relative velocities are evaluated assuming all planetary orbits to be circular, in the same plane as the orbit of the Earth. In this table, relative velocities are positive if the spacecraft goes faster than the planet or asteroid and negative if the spacecraft goes slower than the planet or asteroid. The departure and arrival conditions further assume that the Earth and the target body are at the appropriate relative positions in the solar system.

velocity in this optimistic hypothesis is not far from the upper limit in Figure 1. Direct transfer missions to Mercury, Jupiter, Saturn, and even more so Uranus and Neptune are out of reach of medium-class launchers such as Soyuz if any significant mass is to be brought to destination. Another important conclusion of this preliminary analysis is that missions to Mars and Venus require relatively short transfer times, while missions to the outer solar system require relatively long cruise durations, apart from being energetically more demanding.

10.22.2.3 Rendezvous Missions

For a rendezvous mission, the spacecraft has to be captured in orbit around the target body. The simplest way to get into orbit is to brake closest to the target body using the spacecraft thruster. This is the exact counterpart of the launch strategy from a highly elliptical orbit, which was considered in Section 10.22.2.1. The same issues arise on the length of thrust arcs, with a possible impact in terms of gravity losses if one has to thrust at higher altitudes. The closest possible altitude for an orbit insertion maneuver depends on the target body: If it has no atmosphere (Mercury and Vesta), 200 km is a safe distance. If with an atmosphere (Mars and Venus), 300 km is a reasonable choice. If there are rings (Saturn) or intense radiation belts (Jupiter), larger distances may be required due to safety constraints, for example, 80 000 km for Saturn so as to be outside of the rings (one can get closer by flying above the rings, as was done by Cassini). At a given distance from a planet, the local escape velocity V_{esc} defines the orbits, which have a total energy of 0 in the reference frame of the primary. The relationship between the ΔV required for a marginal capture and the arrival velocity V_{arr} can then be derived from energy conservation:

$$(V_{esc} + \Delta V)^2 = V_{esc}^2 + V_{arr}^2 \quad \Delta V = \left(V_{esc}^2 + V_{arr}^2\right)^{0.5} - V_{esc} \qquad [7]$$

From eqn [7], the required ΔV increases with the arrival velocity and decreases with the local escape velocity; hence, it is easier to capture around a large solar system body. The actual orbit insertion maneuver is not much larger than that required for a marginal capture, as the first orbit around the target has always a much longer period than the operational orbit. The 8 days' initial orbit of Mars Express costs 60 m s^{-1} more than a marginal capture around Mars, and the 115 days' initial orbit of Cassini around Saturn costs 140 m s^{-1} more than a marginal capture around Saturn with the same approach velocity.

From Table 3, the large gravity fields of Jupiter and Saturn compensate for the relatively large arrival velocity (Table 2), so that the capture ΔV is in the same range of that required for Venus and Mars (0.34–0.68 km s^{-1}). Conversely, orbiter missions around Mercury or an asteroid such as Vesta require extremely large orbit insertion maneuvers. Once again, it is useful to go back to the rocket equation [3]: with an exhaust velocity of 3.1 km s^{-1}, the mass in orbit can be larger than 80% of the mass during approach for 'easy' rendezvous targets such as Venus, Mars, Jupiter, and Saturn. Conversely, orbiter missions to Mercury and Vesta are out of reach of a single-stage onboard propulsion system unless the relative velocity can be reduced to more reasonable values using gravity-assist strategies (see Section 10.22.6).

Table 3 Capture ΔV requirements for a rendezvous mission

Target	*Arrival (km s^{-1})*	*Local escape velocity (km s^{-1})*	*Capture ΔV (km s^{-1})*
Mercury	9.61	4.09	6.35
Venus	2.71	10.36	0.35
Mars	−2.65	4.81	0.68
Vesta	−4.43	0.3	4.14
Jupiter	−5.64	35.0	0.45
Saturn	−5.44	24.0	0.61

10.22.2.3.1 Rendezvous missions to natural satellites of giant planets

Among the 15 largest solar system bodies (excluding the Kuiper Belt), 7 are in orbit around a planet: the Moon and the six largest natural satellites of giant planets (Io, Europa, Ganymede, and Callisto around Jupiter, Titan around Saturn, and Triton around Neptune). It has already been indicated that a flyby mission to a Galilean satellite has very similar requirements to a flyby mission to Jupiter, and the Voyager 1 and 2 missions did combine these two objectives. The situation is drastically different for orbiter missions: the large gravity fields of giant planets, which make it relatively easy to capture into a highly elliptical orbit, result in very large orbital velocities and local escape velocities at the distances of large natural satellites. As an example, the orbital velocity of Europa is 13.7 km s^{-1}, and the local escape velocity is 19.4 km s^{-1}. Capturing a spacecraft in an orbit with its pericenter at the distance of Europa requires a minimum ΔV of 0.8 km s^{-1} (one can do somewhat better by capturing at a lower altitude and then raising the pericenter, but this extends the mission duration). Matching the orbit of Europa for a rendezvous would require an additional 5.7 km s^{-1}. Therefore, orbiter missions to satellites of giant planets cannot be achieved simply with a chemical propulsion system. Much more complex methods must be implemented, such as satellite gravity-assist methods (Longman and Schneide, 1970) or nuclear electric propulsion (NEP), or a combination of both (see Sections 10.22.6 and 10.22.7).

Apart from Mercury, all planets have a substantial atmosphere. Among natural satellites, this is the case only for Titan. It is then possible to dip into the atmosphere so as to lose energy and lower the orbit (French and Cruz, 1980). This has to be done in one pass if one wants to use the atmosphere for the insertion maneuver (aerocapture) or for a landing module (atmospheric entry). Once in orbit, it is possible to lower the orbital period by dipping into the outer regions of the atmosphere during many successive orbits (aerobraking). In situ missions to planetary bodies with an atmosphere have a close relationship with flyby missions, as an entry module with a heat shield can quite effectively dissipate the arrival energy for a wide range of arrival velocities.

10.22.2.3.2 Rendezvous missions to comets and Earth-crossing asteroids

Comets and Earth-crossing asteroids are very specific as flyby missions and rendezvous missions have completely different characteristics, contrarily to solar system bodies in nearly circular orbits. As has already been mentioned, flyby missions can

be very cheap if the encounter location is not too far from the orbit of the Earth. Unfortunately, such encounters occur at a very high relative velocity (more than 70 km s^{-1} for the flybys of Halley's comet in 1986), which makes it impossible to consider a rendezvous. Two conditions must be met for a rendezvous mission to be feasible with chemical propulsion:

- The inclination must be relatively small ($\leq 12°$), which eliminates most candidates.
- The perihelion must range between 0.65 and 1.5 AU.

Asteroid 433 Eros, the largest near-Earth asteroid, meets these criteria with a perihelion at 1.13 AU and an inclination of 10.8°. It was studied by the Near Earth Asteroid Rendezvous (NEAR) mission of NASA from rendezvous in February 2000 to a programmed touchdown on Eros in February 2001 (Cheng et al., 1997). Comet Churyumov–Gerasimenko, the target of the Rosetta rendezvous mission, is a typical good candidate among comets, with an inclination of 7.1° and a perihelion at 1.29 AU.

The lowest relative velocity at arrival is achieved by departing from the Earth close to the perihelion of the target body, hence arriving close to aphelion. Again, considering the case of Churyumov–Gerasimenko, the minimum arrival velocity is a reasonable 1.2 km s^{-1}, but the departure velocity is close to 10 km s^{-1}, which is out of reach for a direct launch. Similarly to giant planets, gravity-assist strategies (Section 10.22.6) or an advanced propulsion system (Section 10.22.7) is required.

10.22.3 Remote-Sensing and In Situ Missions to Venus and Mars

The simplified analysis in Section 10.22.2 demonstrated that Venus and Mars stand out as the easiest targets for a rendezvous mission apart from the Moon. They are also surrounded by an atmosphere, which is a major asset for implementing in situ modules (landers, rovers, balloons, etc.). It is therefore no surprise that until 1972, all planetary missions were sent toward one of these two planets. However, not all opportunities are equal, due to the nonnegligible inclination of Venus (3.39°) and Mars (1.85°) and the significant eccentricity of the orbit of Mars ($e = 0.093$).

The first issue to consider is how often can one send a spacecraft to Venus and Mars. Arrival occurs nearly 180° away from departure. During the required cruise time from Earth to Venus on a Hohmann transfer trajectory (0.399 years), Venus travels 233°. Therefore, the appropriate departure time is when Venus is 53° behind the Earth on its orbit. For Mars, the cruise time is 0.709 years, during which Mars travels 135.6°. The appropriate departure time is when Mars is 44.4° ahead of the Earth on its orbit (this assumes that the orbit of Mars is nearly circular, which is not the case). How often does such a favorable configuration repeat? This period P_{syn} (called the synodic period) is such that the angle traveled by the innermost planet (which moves fastest) exceeds by one full turn the angle traveled by the outermost planet. If P_1 and P_2 are the periods of the innermost and outermost planets, we have

$$P_{syn}/P_1 = P_{syn}/P_2 + 1 \quad P_{syn} = P_1P_2/(P_2 - P_1) \qquad [8]$$

If one considers the orbital period of Venus (0.615 years) and Mars (1.881 years), this gives a repeat interval of 1.6 years for transfers to Venus and 2.135 years for transfers to Mars. The position of the Earth at launch moves forward by 216° for Venus and by 48.6° for Mars. After 8 years (5 synodic periods, 13 Venus 'years'), the windows to Venus repeat almost perfectly, as Venus is then only 1.5° ahead of its position 8 years before. For Mars, the best match is after 15 years (7 synodic periods), but Mars lags by 9° behind its position 15 years before, and the windows do not repeat. Intermediate launch dates can be obtained by inserting 'phasing' orbits in the transfer, so that the spacecraft travels close to 540° instead of 180°. This increases the cruise time by 0.8 years for Venus and 1.4 years for Mars. A 540° transfer to Venus requires launching 6 months earlier than the direct window. Such opportunities can be useful, in particular when Venus or Mars is used for a gravity assist (see Section 10.22.6).

The significant eccentricity of the orbit of Mars (0.093) has an impact on transfer orbits. The minimum velocity at infinity required to reach Mars is 3.5 km s^{-1} when it is reached at aphelion (1.665 AU) and only 2.3 km s^{-1} when it is reached at perihelion (1.38 AU). Conversely, the minimum arrival velocity is much higher at perihelion (3.3 km s^{-1}) than at aphelion (2.1 km s^{-1}). The range of departure velocities corresponds to a difference of only 300 m s^{-1} of ΔV from the upper stage. The range of arrival velocities corresponds to a difference of 640 m s^{-1} for the insertion maneuver (actually somewhat larger as gravity losses also increase), which is quite significant (20% less mass budget for an orbiter with an insertion at perihelion, if one does not use aerocapture or aerobraking). Venus has a nearly circular orbit, so that in this respect, all Venus windows should be equivalent. The Earth itself has a small eccentricity (0.0167), so that an outbound launch in January (when the Earth is closest to the Sun) and an inbound launch in July require slightly smaller departure velocities.

However, a much more important problem results from the small but significant inclination of the orbit of Mars (1.83°) and even more so Venus (3.39°) with respect to the orbit plane of the Earth (the ecliptic). The plane of the transfer orbit contains the Sun, the Earth at departure, and the target planet at arrival. The planet at arrival is in the ecliptic only at one of the two intersections of the orbit plane and the ecliptic (nodes): $L_s \sim 76.8°$ and $L_s \sim 256.8°$ for Venus and $L_s \sim 49.5°$ and $L_s \sim 229.5$ for Mars (L_s: heliocentric longitude, measured from the Sun–Earth direction at the fall equinox). In most cases, this forbids a perfect Hohmann transfer, as a 180° orbit would have to fly directly over the Sun. For Mars, angles of 160° (type I transfers) or 200° (type II transfers) bring the transfer orbit plane much closer to the ecliptic. The in-plane departure and arrival velocities are suboptimal compared to the Hohmann transfer, but the out of ecliptic component of the departure velocity drops from the staggering values (>30 km s^{-1}) required for a 180° transfer to 2 km s^{-1} or less (when Mars at arrival is close to one of the nodes).

As a conclusion of this analysis, a good window to Mars (Venus) occurs when Mars (Venus) can be reached close to one of the nodes. For Venus, both nodes are equivalent, as they are at nearly the same distance from the Sun (0.7205 and 0.7261 AU for L_s 76.8° and L_s 256.8°, respectively), but

only the latter lies conveniently close to one of the five possible arrival locations in the 8 years' cycle. For Mars, arriving close to L_s 49.5° (at 1.4715 AU) is slightly better than arriving close to L_s 229.5° (at 1.5515 AU) for flyby and in situ missions and slightly worse for orbiter missions relying only on chemical propulsion.

The situation is further complicated as the performances of Ariane 5 from Kourou strongly depends on the declination (angle of the departure velocity to the equatorial plane of the Earth). At the low latitude of Kourou (5° N), the launcher can take full advantage of the rotational velocity of the Earth (461 m s^{-1} instead of 463 m s^{-1} at the equator) by launching due east if the departure velocity lies close to the equatorial plane. For larger declinations, one has to launch away from the optimum direction (avoiding Brazil for the fall of the boosters), and the thrust must be applied at an angle to the velocity ('dogleg') as the third stage of Ariane 5 cannot be restarted. When the launch pad is at a higher latitude (Kennedy Space Center, 28.5° N and Baikonur, 46° N), the rotational velocity is lower (407 m s^{-1} for Kennedy, 322 m s^{-1} for Baikonur), but the launch capability is the same for all declinations (N or S), which do not exceed the latitude. The spacecraft is first set on a circular orbit with a declination equal to the latitude by launching due east (which is optimum). The upper stage (Centaur, Fregat) is then restarted at the position on the circular orbit, which is adequate for reaching the departure declination. Soyuz–Fregat can be launched from Kourou since late 2011. The circular orbit has then a declination equal to that of departure, with a much smaller mass penalty than for Ariane 5 (the same as for a launch at the corresponding latitude). As there is an angle of 23.45° between the ecliptic and the equator, the declination is a complex combination of the departure date and the out of ecliptic component.

Tables 4 and 5 provide useful comparisons for all remote-sensing and in situ missions to Mars and Venus. They show the major impact of the 3.39° inclination of Venus on the approach velocity, which never gets close to the lower limit derived from the coplanar assumption. Conversely, several Mars windows have overall performances that are very close to that of the coplanar assumption. The departure or arrival velocity for Mars can even be smaller than the coplanar estimate, as Mars (0.093) and the Earth itself (0.0167) have non-circular orbits. The similarity between the Venus Express transfer orbit (2005/08) and that of the 2013 window illustrates the 8 years' cycle of windows to Venus (5 synodic periods of 1.6 years). With Soyuz from Baikonur (Mars Express), many windows are relatively similar for an orbiter mission. With Ariane 5, specific strategies must be considered for all windows from 2009 to 2020 due to declination (25° N or 46° S is not accessible). Inserting a 1-year orbit before departure or using a favorable type III window (2017) results in a cruise time of 1.7–2.2 years.

10.22.3.1 Orbiter Missions Not Implementing Aerocapture

Orbiter missions that do not implement aerocapture have to use their propulsion system to insert into orbit around the planet. The simplest way to compare windows is to consider as a baseline for all missions the launch strategy from a Moon-crossing orbit described in Section 10.22.2.1, which provides the best mass budget on Mars and Venus transfer trajectories. The spacecraft propulsion system is used both to leave Earth orbit and for orbit insertion. Therefore, high departure or arrival velocities have a comparable impact in terms of total ΔV (next to last column of Tables 4 and 5). Assuming a 3.1 km s^{-1} ejection velocity for the spacecraft thruster, this can be translated into an available mass in orbit by applying the rocket equation [3], to the launch capability in a large bound orbit for a given launcher (Soyuz–Fregat for the last column in Tables 4 and 5).

After capture in a long-period orbit (8 days for Mars Express), it is possible to reduce the period of the orbit by implementing thrust at pericenter. 2.96 km s^{-1} of additional ΔV is required to go down to a circular orbit around Venus at an altitude of 300 km, and 1.41 km s^{-1} of additional ΔV is required to reach a circular orbit around Mars at an altitude of 300 km. Splitting the maneuver over many pericenters (eight for Mars Express) limits gravity losses to a few m s^{-1}, as each thrust arc is short, close to the optimum altitude.

From Table 4, the 2005 window to Venus was attractive in terms of departure velocity, but it was relatively poor in terms of arrival velocity. This was not a major issue as the orbital period could not be shorter than 1 day due to constraints on eclipse duration imposed by the reuse of the Mars Express spacecraft.

From Table 5, the 2003 window to Mars stands out as one of the best over 20 years. With a direct injection by Soyuz–Fregat, the Mars Express spacecraft (1050 kg after releasing the 70 kg

Table 4 Launch windows to Venus for Venus Express (10/2005) from 2013 to 2018

Launch	V_{dep} (km s^{-1})	*Launch declination*	*ΔV/Moon*	*Arrival position*	*Arrival date*	V_{arr} (km s^{-1})	*ΔV/V_{esc}*	*ΔV, total*	*Mass (kg)*
Coplanar approximation	2.500		0.338			2.710	0.357	0.695	1287
2005/10/31 (II)	2.780	26.4° S	0.404	L_s 249°	2006/04/07	4.653	1.019	1.423	1017
2013/10/30 (II)	2.795	26.6° S	0.408	L_s 250°	2014/04/06	4.730	1.052	1.459	1006
2015/05/21 (II)	2.480	9.7° N	0.333	L_s 78°	2015/10/27	3.780	0.683	1.017	1160
2016/12/28 (I)	2.664	3.0° N	0.376	L_s 257°	2017/05/09	3.754	0.674	1.050	1147
2018/06/15 (II)	3.891	−42.2° S	0.728	L_s 116°	2018/12/17	3.061	0.453	1.182	1100
2018/08/22 (I)	3.107	9.1° N	0.489	L_s 102°	2018/12/08	4.562	0.981	1.470	1002

The type of transfer (I or II) is indicated. The ΔV at departure is evaluated assuming an indirect launch from a Moon-crossing orbit. The ΔV at arrival is evaluated for a marginal capture at an altitude of 300 km. The mass in orbit is evaluated for a Venus Express-type launch (Soyuz–Fregat from Baikonur: no penalty for declination up to 46° N or S).

Table 5 Launch windows to Mars for Mars Express (06/2003) from 2005 to 2022, with the same definitions as for Table 4

Launch	V_{dep} (km s^{-1})	Launch declination	ΔV/Moon	Arrival position	Arrival date	V_{arr} (km s^{-1})	$\Delta V/V_{esc}$	ΔV, total	Mass (kg)
Coplanar approximation	2.940		0.444			2.650	0.681	1.125	1120
2003/06/03 (I)	2.990	3.9° S	0.458	L_s 50°	2003/12/26	2.693	0.702	1.159	1108
2005/08/23 (II)	4.000	0.3° N	0.766	L_s 183°	2006/08/26	3.102	0.912	1.678	937
2007/09/14 (II)	3.638	9.1° N	0.646	L_s 203°	2008/08/26	2.531	0.625	1.270	1069
2009/10/14 (II)	3.200	20.6° N	0.515	L_s 228°	2010/09/03	2.465	0.594	1.109	1126
2011/11/11 (II)	3.000	32.3° N	0.460	L_s 258°	2012/09/17	2.720	0.715	1.175	1102
2013/12/06 (II)	3.075	24.5° N	0.480	L_s 289°	2014/09/27	3.180	0.955	1.435	1013
2016/03/07 (I)	3.355	46.1° S	0.559	L_s 314°	2016/09/24	3.855	1.353	1.912	869
2017/10/29 (III)	3.320	0°	0.550	L_s 222°	2020/01/16	3.160	0.940	1.490	996
2018/05/13 (I)	2.800	35.2° S	0.409	L_s 26°	2018/12/05	2.945	0.829	1.238	1080
2020/07/25 (I)	3.700	26.2° N	0.667	L_s 88°	2021/02/13	2.650	0.681	1.358	1039
2022/09/05 (II)	3.880	0°	0.726	L_s 195°	2022/08/26	2.650	0.681	1.407	1023

Beagle II module) could brake down to a 6.7 h polar orbit, which required 1.5 km s^{-1} of ΔV including a plane turn, leaving 70 kg of reserve fuel in the tanks on top of the 570 kg dry mass. The 2005 window is among those delivering the lowest mass in Mars orbit (20% less than in 2003, representing, e.g., 170 kg of fuel). This was one of the arguments against flying a new mission to Mars as a follow-up to Mars Express, selecting instead Venus Express, successfully launched in October 2005. The exploration program of ESA was initially considering a rover mission in the 2013–15 time frame (it has since then been significantly delayed). As the present design is based on a large rover ('ExoMars'), this mission belongs to the category of in situ missions, which will be considered next. Orbiter missions may be required after 2010 so as to provide relay capability. In this respect, 2013 is poor and 2016 is bad, and the next good window after 2011 is in 2018.

The mass budget of orbiter missions can be markedly improved by aerobraking. This method involves shallow atmospheric passes, which slowly reduce the semimajor axis starting from a highly elliptical capture orbit. The impact on spacecraft design is significant, but far less severe than for aerocapture. Solar panels pick up most of the strain. They have to be symmetrical and relatively robust, but no heat shield is required. This method was first implemented for the Mars Global Surveyor (MGS) mission of 1998 (Albee et al., 2001), then for Mars Odyssey in 2000 (Smith and Bell, 2005), and in late 2006 for the MRO mission, which was launched in August 2005. Aerobraking was definitely required for MRO as it is a large spacecraft, while the 2005 window is poor. If one goes all the way down to a circular orbit, which is the case for both MGS and MRO, nearly 1.4 km s^{-1} of ΔV is saved, which corresponds to a factor of 1.57 in terms of mass budget (actually closer to 1.7 if one takes into account the additional mass of tanks and structure). There is a drawback: as each pass has to be gentle, the orbit reduction takes a long time. The science phase of the mission was therefore delayed by 6 months for MRO, and it was delayed by 1 year for MGS, due to a nonnominal configuration of one of the solar panels. There is some risk involved, and the aerobraking phase requires a high level of monitoring, with a significant impact on operational cost. Aerobraking has yet to be implemented by ESA, but it should definitely be considered for large orbiters in low orbits around Mars or Venus.

10.22.3.2 In Situ Missions and Orbiter Missions Implementing Aerocapture

For in situ missions (landers, rovers, balloons, etc.) or orbiter missions implementing aerocapture, an entry module with a heat shield dissipates the approach energy. The control of a trajectory deep in an atmosphere is a complex problem, which is beyond the scope of this chapter. It is extensively discussed for Earth reentry trajectories (see, e.g., Vinh and Kuo, 1997; Willcockson, 1999; Wu et al., 2001) and for other planetary bodies (see, e.g., Justus et al., 2005; Scoon et al., 1991; Tu et al., 2000. There is a narrow entry corridor for aerocapture: If the minimum altitude is too high, the velocity remains above escape, and the spacecraft leaves the planet. If the minimum altitude is too low, the velocity drops below the circular velocity, and the spacecraft falls on the planet. The latter case is what is required for an in situ mission. However, if the entry is too steep, the heat flux may overcome the dissipation capability of the shield, and the mission is lost. The energy corresponding to the escape velocity at the surface (10.3 km s^{-1} for Venus and 5.1 km s^{-1} for Mars) has to be dissipated anyway; hence, the mass that is required for the heat shield depends only weakly on the approach velocity: the range of specific energies to be dissipated is 57–64 MJ kg^{-1} for Venus (see, e.g., Craig and Lyne, 2005) and 16.5–20.5 MJ kg^{-1} for Mars for the full range of approach velocities in Tables 4 and 5. Therefore, for such missions, the most important parameters are the departure performances, which can be defined from columns 3 (declination) and 4 (ΔV at departure from a bound orbit) of these tables. For such types of missions, a window with a low departure velocity and a high arrival velocity becomes attractive, and the type I window to Venus in 2018 would be selected (3.1 km s^{-1} at departure) instead of the type II window (3.7 km s^{-1} at departure), while the latter would provide more mass for an orbiter with chemical insertion. ESA has

demonstrated its capability to deliver a surface station to Titan with the Huygens mission, with an entry velocity of 6 km s^{-1} very similar to Mars entry velocities. Typically, half of the mass of an entry module is available as useful mass on the surface, as demonstrated by missions such as Huygens (ESA) or the 2004 Mars rovers (NASA). Launchers such as Atlas/Centaur (NASA) or Ariane 5 (ESA) can deliver 4000 kg to Mars, making it possible to deploy a large surface module. The Mars Science Laboratory (MSL) mission of NASA, launched in 2011, is an example of such ambitious undertakings with its 800 kg 'Curiosity' rover. A combination of a large relay orbiter (1 t) and test descent module launched in 2016 and a larger entry and descent module launched in 2018 is now the baseline for the ExoMars mission of ESA's exploration program, to be implemented in collaboration with Russia. Similarly to Venus and Mars, aerocapture provides the most effective strategy for other planetary bodies with a significant atmosphere. It is in particular likely that a post-Cassini mission to Titan will have to implement such a strategy (e.g., Spilker, 2005).

10.22.4 Orbit Evolution Around Planetary Bodies

For all orbiter missions, the evolution of the orbit with time is a critical aspect of science operations, as it defines which regions can be observed and the local time (a critical parameter for cameras and near-IR spectrometers, which rely on solar photons).

The perturbations of the orbit by the Sun play only a minor role except for highly eccentric orbits. The main effect is a change in the minimum altitude, which has to be compensated if it decreases so as to avoid impacting the planet. When approaching a planet at a low velocity, the combined gravitational pull of the Sun and the planet can provide for free a capture in a weakly bound orbit. These three body effects are important for planetary systems, in particular for the Earth–Moon system (Belbruno and Miller, 1993; Knezevic and Milani, 1998).

High-order terms of the gravity field of the planetary body are most important for orbits closest to the target body. They result in a precession of the intersection of the orbit plane with the equatorial plane (defined by the position of the ascending node Ω) and a precession of the pericenter (defined by the angular distance from the ascending node to the pericenter ω) for elliptical orbits. The quadrupolar term of the gravity field, J_2, plays the most important role. It is linked to the oblateness of the planetary body. The evolution of Ω and ω as a function of J_2 and orbital elements (see, e.g., Hechler et al., 2005) is the following:

$$p = a\left(1-e^2\right) \quad [9]$$

$$\dot{\Omega} = -\frac{3}{2} n J_2 \left(\frac{R_m}{p}\right)^2 \cos i$$
$$\dot{\omega} = \frac{3}{4} n J_2 \left(\frac{R_m}{p}\right)^2 \left(4 - 5\sin^2 i\right) \quad [10]$$

where R_m is the radius of the planetary body, a is the semimajor axis, e is the eccentricity, n is the mean motion of the spacecraft on its orbit (inversely proportional to $a^{1.5}$), and i is the inclination. For circular orbits, the parameter (p in eqn [9]) is equal to a, and the precession of the plane is inversely proportional to the semimajor axis$^{3.5}$.

A specific case of interest is the Sun-synchronous orbit, with a full rotation of the orbit plane in exactly one orbital period of the planetary body around the Sun. With such an orbit, the local time remains approximately constant. This is a desirable feature for cameras and IR imaging spectrometers as observations on successive orbits are made with the same illumination conditions. With the value of J_2 for the Earth (0.001082), an inclination of 97.8° is needed for a circular orbit at 600 km altitude from eqn [9]. A Sun-synchronous orbit has been selected for NASA missions orbiting Mars on a low circular orbit, such as MGS and MRO. The J_2 value for Mars (0.001965) is larger than that of the Earth and the orbital period of Mars is 687 days, so that a lower value of $\cos(i)$ is required for Sun-synchronous orbits. The inclination of MGS (altitude ~400 km) is 92.9° and that of MRO, which reached a circular orbit at an altitude of ~300 km in November 2006, is 92.7°. MRO crosses the equator in mid-afternoon (3 PM) on average. However, the rate of precession is constant, while the angular motion of Mars is 45% larger at perihelion (1.382 AU) than at aphelion (1.666 AU) due to Kepler's 2nd law. As a consequence, the local time at the equator varies from 2 h15 to 3 h 45 PM during one Martian year. The variations in local time are only 2 min for Sun-synchronous orbits around the Earth due to the low eccentricity of its orbit (0.0167).

The precession of the pericenter of an elliptical orbit around Mars such as that of Mars Express is defined by eqn [10]. This precession provides opportunities for observing each latitude range at the highest possible resolution. However, combining the rotation of the planet around the Sun, the precession of the orbit plane and the precession of the pericenter so as to optimize the science return are a complex problem (Hechler et al., 2005), in particular when both dayside observations (for imaging experiments) and nightside observations (for the MARSIS radar) are required at low altitude. The orbit evolution of the selected orbit is presented in **Figure 2**. It provides deep excursions of the pericenter over the sunlit side (L_s 135°, L_s 315°, and L_s 90° in 2006) and shallow excursions of the pericenter beyond the terminator (L_s 90° and L_s 225°).

Venus has a very small oblateness, which can be attributed to its very slow rotation rate. Therefore, precession was neither an issue nor an asset for orbiter missions to Venus such as Magellan (NASA) or Venus Express (ESA). The pericenter of the highly elliptical orbit of Venus Express will therefore remain at high northern latitudes for the full science mission. The value of J_2 for Mercury as measured by Messenger (NASA) from orbit is significantly higher than that inferred from the Mariner 10 flybys. The resulting precession is negligible for the highly eccentric orbit of Messenger of that of the Mercury Magnetospheric Orbiter of BepiColombo (ESA/JAXA). Due to this revised evaluation of the gravity field, the mission profile of the Mercury Planetary Orbiter of BepiColombo, which has a low eccentricity (0.16), had to be redesigned so as to optimize observation conditions and to prevent collision with the planet.

10.22.5 In Situ Missions to Atmosphereless Bodies

In situ missions to atmosphereless bodies represent a specific challenge, as one cannot rely on an atmospheric entry module. The only solution for landing safely on the surface consists in

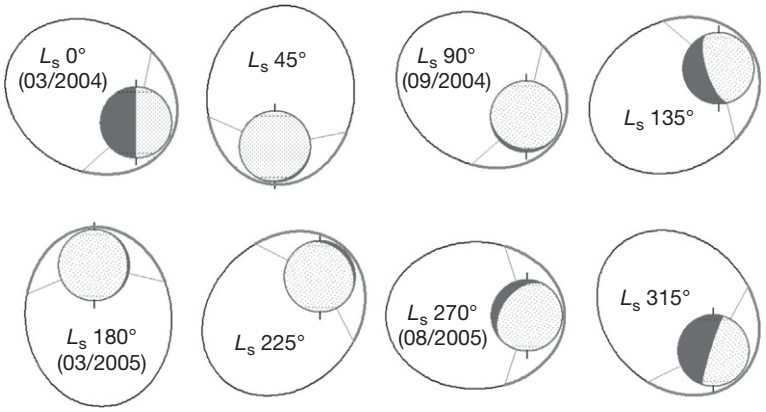


Figure 2 Evolution of the orbit of Mars Express during the first Martian year of science operations (nominal mission). For each heliocentric longitude L_s, Mars is shown as seen from a direction perpendicular to the orbit, the dotted regions being sunlit. The pericenter rotated 1.22 times around the planet during one Martian year, so that each band of latitude was observed at least twice at the lowest altitude (~280 km) at different local times.

Table 6 Escape velocity and surface gravity of large atmosphereless bodies

	Mercury	*Moon*	*Io*	*Europa*	*Ganymede*	*Callisto*	*Vesta*	*Ceres*
Size (km)	4880	3470	3640	3120	5280	4820	540	945
V_{esc} (km s^{-1})	4.257	2.376	2.560	2.030	2.741	2.440	0.370	0.515
g (m s^{-2})	3.700	1.622	1.800	1.320	1.430	1.235	0.260	0.280

using an onboard propulsion system. A soft landing is in many ways similar to a launch in reverse. There is one piece of good news: contrary to a launch, the mass decreases as the landing module gets near the surface; hence, the acceleration from the thrusters increases, while the acceleration from gravity remains nearly constant. Again, similarly to a launch, a landing requires a ΔV that is larger than the escape velocity (or the orbital velocity at pericenter if the lander is released from orbit). The propulsion system has to land itself. A large thrust level requires more mass for thrusters. Conversely, if the thrust is low, there are gravity losses and a larger fraction of the thrust is wasted in the final stages fighting the local gravity.

Atmosphereless solar system bodies of interest are comets, asteroids, KBOs (among them Pluto and probably Triton), small natural satellites, six of the seven largest natural satellites (the Moon, Io, Europa, Ganymede, Callisto, and Triton), and one planet (Mercury). Landing on a cometary nucleus a few km across is a difficult navigation problem, but it is not a mass budget issue, as escape velocities are only in the range of meters per second. There is no survey mission planned yet for Pluto or Triton; therefore, an in situ mission to KBOs is far in the future. The most relevant issues are therefore that of landing on Mercury, the Moon, Galilean satellites, and large asteroids such as Vesta.

Table 6 shows the escape velocity and surface gravity of relevant large atmosphereless bodies. From this table, it is clear that in terms of total ΔV, there is not much difference between a soft lander (a few m s^{-1}), a semihard lander with airbags (~50 m s^{-1}), and a penetrator (up to 100 m s^{-1}), with the exception of asteroids such as Vesta or Ceres.

From Table 6, Mercury, while not the largest atmosphereless body, is in a class by itself in terms of landing: The surface gravity is more than twice that of any other atmosphereless body, and the energy to be dissipated is 2.35–4 times larger than that required for Galilean satellites. The mass budget required for a lander is therefore very large.

With its ion propulsion stage (see Section 10.22.7), BepiColombo approaches Mercury on a weakly bound orbit. The chemical propulsion stage sets the composite into the orbit of the eccentric orbiter (MMO). A release of a landing module was considered at this stage during the early study phases. The key issue is the fraction of the mass of the landing module, which can reach the surface, and the smaller fraction, which is left for lander systems and payload ('useful mass'). A simple landing strategy, close enough to optimum for this order of magnitude analysis, consists in tilting the thrust toward the center of the planet so that the vertical component balances gravity together with the remaining centrifugal force. The remaining horizontal component is used to reduce the orbital velocity. The critical parameter is the acceleration level at approach. Table 7 indicates the thrust time, total ΔV (evaluated from approach), and mass breakdown of a landing module using this strategy with initial acceleration levels between 0.75 and 2 m s^{-2}.

As expected, a larger thrust level provides a larger landing mass as less time is spent fighting the 3.7–3.8 m s^{-2} of local gravity. A useful mass of 50 kg for 5–10 kg of science payload can be considered as a minimum. With an initial acceleration level of 1.5 m s^{-2}, the useful mass has to be multiplied by ~16 (1/0.064) to obtain the initial mass. Therefore, a minimum mass of 800 kg (margins excluded) is required for the landing module and the corresponding fuel for orbit insertion. The results of this simplified analysis are fully consistent with that of the ESA study for the BepiColombo mission to

Table 7 Elements of a mass budget trade-off for a Mercury lander resulting from the simple landing strategy presented in the text

Initial acceleration	0.75 m s^{-2}	1 m s^{-2}	1.5 m s^{-2}	2 m s^{-2}	∞
Thrust time (s)	3660	2640	1650	1200	0
Total ΔV (km s^{-1})	6.70	5.88	4.99	4.62	4.24
Mass on the surface	0.115	0.150	0.200	0.227	0.255
Propulsion system	0.028	0.038	0.056	0.075	NA
Tanks and structure	0.089	0.085	0.080	0.077	NA
Useful mass	−0.002	0.026	0.064	0.075	NA

The total mass on the surface, the mass of the propulsion system, the mass of tanks and structure, and the useful mass are expressed as a fraction of the mass of the landing module at approach.

Mercury. The large mass allocation required was a key element in the decision to descope the landing module in mid-2003.

Other large atmosphereless bodies provide much more favorable constraints for a lander. It turns out that a 1 m s^{-2} acceleration level at approach is adequate for all these bodies. Table 8 provides the minimum mass of a landing module for a 50 kg smart lander, which controls its trajectory to the ground.

The important role of surface gravity is demonstrated by the slightly more favorable mass budget for Ganymede than for Io, while its escape velocity is significantly larger (2.74 km s^{-1} instead of 2.65 km s^{-1}, Table 6). Even with more conservative assumptions, it is clear that a landing module on these bodies (in particular Europa) can be considered for a mass budget not exceeding 200–250 kg at approach. This mass budget has of course to be multiplied by the proper factor from the rocket equation [3], to account for ΔV expenditures from Earth departure to approach at a low velocity.

Table 8 Mass of the landing module on a low-velocity approach to four medium-sized atmosphereless bodies of interest, which is needed to implement a smart lander with a useful mass of 50 kg, assuming a 1 m s^{-2} acceleration level at approach on a weakly bound orbit

	Moon	*Io*	*Europa*	*Ganymede*
Impulsional ΔV (km s^{-1})	2.37	2.57	2.03	2.75
Actual ΔV (km s^{-1})	2.93	3.17	2.40	3.12
Mass on the surface	0.387	0.359	0.460	0.365
Propulsion system	0.038	0.038	0.038	0.038
Tanks and structure	0.061	0.064	0.054	0.063
Useful mass	0.289	0.257	0.368	0.264
Mass of landing module	173 kg	194 kg	136 kg	189 kg

Similarly to **Table 7**, masses are expressed as a fraction of the initial mass on such an orbit.

10.22.6 Gravity-Assist Missions: Giant Planets, Mercury, Asteroid, and Comet Rendezvous

As demonstrated in Section 10.22.2.1, flyby missions to giant planets and Mercury and rendezvous missions to comets require departure velocities higher than 7 km s^{-1}, which are difficult to reach with large useful masses by a direct launch or an indirect launch strategy. A wide range of missions to these bodies became possible by using gravity-assist strategies.

A planetary gravity assist consists in passing close to a planet so as to change the trajectory of a spacecraft. In the planetary referential, energy must be conserved; hence, the relative velocity at departure is the same as the relative velocity at arrival. However, the planetary encounter can rotate this relative velocity by an angle that increases with the mass of the planet and decreases with the distance at closest approach. The benefit appears in the heliocentric referential: passing behind a planet significantly increases the heliocentric velocity (Figure 3). In such a case, the planet has provided additional orbital energy to the spacecraft in the heliocentric referential. Similarly, passing in front of the planet reduces the orbital energy without expenditure of onboard fuel.

The geometry of the swingby is defined by the relative velocity at arrival V_{rel} and the minimum approach distance R_p compared to the radius R of the planetary body and the circular velocity V_c at the surface (a factor of $\sqrt{2}$ smaller than the escape velocity V_{esc}). V_c is directly linked to the gravitational potential of the planet GM. The determination of the semimajor axis (a) and eccentricity (e) of the hyperbolic trajectory around the planet is straightforward from Newton's equations of motion:

$$GM/R^2 = V_c{}^2/R$$
$$a = -GM/V_r^2 = R(V_c/V_{rel})^2$$
$$e = 1 - R_p/a$$

The true anomaly v of the asymptote is defined by

$$1 + e\cos(v) = 0$$

The angle of rotation θ is

$$\theta = 2v - \pi$$

The relationship between the rotation angle θ, the minimum distance R_p, and the relative velocity V_r can then be established, as well as the change in velocity ΔV:

$$\theta = 2a\sin(1/e) = 2a\sin\left(1/\left(1 + R_p/R(V_{rel}/V_c)^2\right)\right) \quad [11]$$

$$\Delta V = 2V_{rel}/\left(1 + R_p/R(V_{rel}/V_c)^2\right) \quad [12]$$

From eqn [12], the maximum theoretical ΔV provided by the planet is equal to V_c. It is obtained for $V_{rel} \sim V_c$ and $R \sim R_p$. From Table 9, any planetary encounter can provide a large ΔV for free. This is also the case to a lesser extent for large natural satellites (Table 6).

The first approach that can be considered is the ΔV gravity assist (ΔVGA), which increases the effectiveness of the available onboard ΔV by a large factor by combining a deep space maneuver with a gravity assist by the Earth (see, e.g., Casalino et al., 1998). ΔVGA strategies can also be used with other

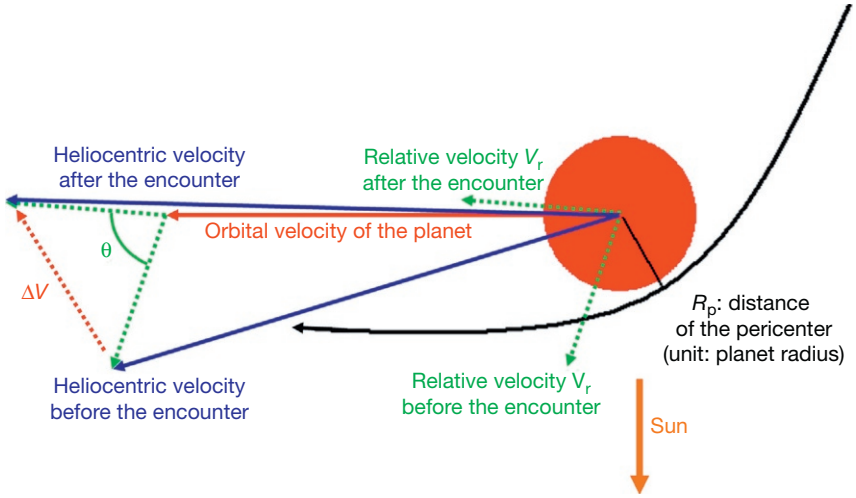


Figure 3 Example of a planetary gravity assist, which increases the orbital energy of a spacecraft in the heliocentric referential. The spacecraft passes over the trailing side. The gravitational pull of the planet rotates the relative velocity at arrival (mostly inbound) in the forward direction at departure, when it adds up to the orbital velocity of the planet. The resulting heliocentric velocity is therefore larger. Such an encounter is included in any ΔVGA strategy.

Table 9 Equatorial radius and circular velocity of planets, corresponding to the orbital velocity on a grazing circular orbit

	Mercury	*Venus*	*Earth*	*Mars*	*Jupiter*	*Saturn*	*Uranus*	*Neptune*
Radius (km)	2440	6050	6380	3400	71 000	60 000	25 500	25 000
V_c (km s^{-1})	3.01	7.33	7.91	3.55	42.1	25.1	15.1	16.6

planets or natural satellites. A ΔVGA strategy can be associated to any resonant orbit that goes back to the Earth after a given number of revolutions. The departure velocity is parallel to the Earth velocity (either forward or backward). The deep space maneuver is performed close to aphelion, reducing the semimajor axis for orbits with periods larger than 1 year, and close to perihelion, increasing the semimajor axis for orbits with periods smaller than 1 year. As a result of this maneuver, the relative velocity when returning to the Earth increases by a much larger amount than the ΔV spent in the maneuver. This is due to the relationship between energy (E) and angular momentum (M) at pericenter ($p = 1$ AU) and apocenter (q). The heliocentric velocity at departure V_p is tangential to the orbit of the Earth. As a function of the eccentricity e, we have for an outbound orbit, following the presentation of Langevin (2005),

$$\begin{aligned} a &= 1/(1-e) \\ q &= (1+e)/(1-e) \\ \Delta M &= q\Delta V = (1+e)/(1-e)\Delta V \\ V_p &= V_e(1+e)^{0.5} = V_e + V_{rel} \end{aligned} \quad [13]$$

where V_e is the heliocentric velocity of the Earth and V_{rel} is the departure velocity:

$$\begin{aligned} V_q &= V_p(1-e)/(1+e) \\ \Delta E_q &= V_q\Delta V = -V_p(1-e)/(1+e)\Delta V \end{aligned} \quad [14]$$

When returning to 1 AU (assuming a resonant orbit, so that the Earth is at the right place), the large reduction in angular momentum at aphelion results in a large reduction of the tangential velocity V_t with respect to its initial value (V_p):

$$V_t = V_p - (1+e)/(1-e)\Delta V$$

As the radial velocity is 0 at departure, if V_r is the radial velocity upon return, we have

$$\Delta E_p = 0.5\left(V_t^2 + V_r^2 - V_p^2\right)$$

For maneuvers that are small with respect to the heliocentric velocities, we can write

$$0.5\left(V_t^2 - V_p^2\right) = -V_p(1+e)/(1-e)\Delta V \quad [15]$$

There is an energy deficit at this stage, when comparing [15] and [14], which must be compensated by the radial component of the velocity V_r. Energy conservation leads to

$$\begin{aligned} \Delta E_p &= \Delta E_q \\ V_r^2 &= 2V_p\Delta V((1+e)/(1-e) - (l-e)/(1+e)) \\ V_r &= \left(8V_p\Delta Ve/(1-e^2)\right)^{0.5} \end{aligned} \quad [16]$$

As an example, for a 3 years' resonant orbit with a departure in January at 6.84 km s^{-1}, the eccentricity is 0.527, and the initial heliocentric velocity V_p is 37.14 km s^{-1}. The radial velocity at return is 2.16 km s^{-1} for a 22 m s^{-1} maneuver at aphelion and 4.60 km s^{-1} for a 100 m s^{-1} maneuver, close to the results of eqn [16]: 2.19 km s^{-1} and 4.67 km s^{-1}, respectively. These small differences can be attributed to the slight eccentricity of the Earth orbit. When evaluating the relative

velocity at Earth return, we have to combine this radial velocity with the tangential relative velocity, which is reduced by the same amount as the tangential heliocentric velocity. As long as V_r remains small with respect to V_{rel}, we have

$$dV_{rel}/\Delta V \sim 4V_p/V_{rel}\, e/\left(1-e^2\right) - (1+e)/(1-e) \qquad [17]$$

The relative velocity at return for a 22 m s^{-1} maneuver is 7.104 km s^{-1}, a velocity increment of 260 m s^{-1}, while eqn [17] would result in a velocity increment of 277 m s^{-1}. Even with such a small maneuver, the radial velocity upon return V_r (2.16 km s^{-1}) is already not that small with respect to V_{rel} (6.84 km s^{-1}). The efficiency of the maneuver decreases as the radial velocity at the Earth increases. Therefore, the ratio between the velocity increment and the ΔV (leverage factor) increases with the eccentricity of the orbit, and it decreases with the velocity gain, which is to be obtained.

A similar derivation for inbound orbits yields

$$V_p = V_e(1-e)^{0.5} = V_e - V_{rel} \qquad [18]$$

$$dV_{rel}/\Delta V \sim 4V_p/V_{rel}e/\left(1-e^2\right) - (1-e)/(1+e) \qquad [19]$$

For the same eccentricity e, an inbound ΔVGA is less favorable than an outbound ΔVGA, as V_p is now smaller than the heliocentric velocity of the Earth [18], contrary to the outbound case [13]. The negative impact of the lower V_p/V_{rel} ratio is not compensated by the reduced value of the second term of [19] compared with that of [17].

Flying over the trailing side of the Earth (for outer solar system missions) or the leading side of the Earth (for inner solar system missions) rotates the relative velocity so that it is nearly aligned with the orbital velocity of the Earth after the encounter (**Figure 3**). This is most effective for reaching distant targets. The increment or decrement in heliocentric velocity over that of the Earth is the effective velocity, as it is equivalent to the departure velocity, which would be provided by a direct launch in the direction of the heliocentric velocity of the Earth (or opposite to it for inner solar system missions). When the relative velocity is very large (9 km s^{-1} or more), a flyby at the lowest acceptable altitude (~250 km) is not able to fully rotate it to the optimum direction (see eqn [11]). The resulting effective velocity is then lower than the relative velocity, and the leverage factor decreases correspondingly. The most relevant examples are provided in **Table 10**. The launch ΔV is that that would have to be spent at the perigee of a Moon-crossing orbit so as to obtain the required departure velocity. For the cheapest orbit, an MGA can provide the required 1.1 km s^{-1} departure velocity (Farquhar, 1991), so that no additional ΔV is required.

The comparison between **Table 10** and **Table 1** demonstrates the potential of gravity-assist methods for the most demanding missions: a 9 km s^{-1} departure velocity required a velocity at perigee 3.29 km s^{-1} higher than that of a Moon-crossing orbit, which could not be achieved by a Fregat upper stage, and only marginally so with a Centaur upper stage. For such a high ΔV, a spacecraft engine burn around perigee (indirect launch strategy) would incur a staggering gravity loss. With a 2:1 ΔVGA, only 1.81 km s^{-1} is needed, with 1.26 km s^{-1} on the bound orbit so as to leave the Earth with a relative velocity of 5.28 km s^{-1} and 0.552 km s^{-1} as a deep space maneuver (for which there is no gravity loss). Using a 2:3 ΔVGA, only 1.71 km s^{-1} is needed to reach the 7.5 km s^{-1} negative velocity required for a direct Mercury flyby mission. With this approach, a spacecraft with a useful mass of about 1000 kg can be sent to Jupiter, a comet, or Mercury with the 2010 version of the Soyuz–Fregat (launched from Kourou). There is a penalty of 2 years in cruise time (hence a 5 years' cruise to Jupiter), which is a small price to pay as such missions are now feasible with classical propulsion methods.

It would seem possible to further improve the mass budget by using a series of ΔVGA maneuvers. However, for departure velocities lower than 4 km s^{-1}, there is not much to be gained from a ΔVGA. The total ΔV budget required for a departure at 3.41 km s^{-1} starting from a Moon-crossing orbit is 0.48 km s^{-1} for a 5:4 ΔVGA and 0.47 km s^{-1} for a 4:3 ΔVGA, a 100 m s^{-1} improvement over the 0.57 km s^{-1} ΔV required for a direct departure from a bound orbit (the gravity loss is small at such low departure velocities). The 3% improvement of the mass budget is definitely not appealing considering the 4–5 years' penalty in cruise time. Using a sequence of a 3:2 ΔVGA followed by a 2:1 ΔVGA for a Jupiter mission instead of a direct 2:1 ΔVGA is more attractive, as the ΔV reduction is now 0.42 km s^{-1} (1.26–0.84 km s^{-1}), actually ~600 m s^{-1} as gravity losses become quite significant for departure velocities in the range of 4 km s^{-1} or more. The improvement of the mass budget is ~20%, still a marginal proposition when considering the 3 years' penalty in terms of cruise time. The 1:1 ΔVGA is specific as the departure velocity can be provided by a Moon

Table 10 Characteristics of relevant ΔV gravity-assist launch strategies

Resonance	*Departure velocity (km s^{-1})*	*Launch ΔV (km s^{-1})*	*Eccentricity*	*Maneuver (km s^{-1})*	*Effective velocity (km s^{-1})*	*Total ΔV (km s^{-1})*	*Leverage factor*	*Additional cruise time (years)*
1:1	−1.100	MGA	0.076	1.000	4.1	1.00	3.00	1
2:3	−5.342	1.286	0.329	0.428	7.55	1.72	5.16	2
3:4	−3.475	0.591	0.220	0.327	5.14	0.92	5.09	3
5:4	+2.020	0.236	0.141	0.246	3.41	0.48	5.64	5
4:3	+2.520	0.337	0.177	0.136	3.41	0.47	6.55	4
4:3	+2.586	0.352	0.181	0.485	5.17	0.84	5.33	4
3:2	+3.412	0.571	0.242	0.265	5.17	0.84	6.63	3
2:1	+5.171	1.214	0.377	0.224	6.98	1.43	8.07	2
2:1	+5.283	1.261	0.386	0.552	9.03	1.81	6.78	2
3:1	+6.983	2.095	0.523	0.195	9.03	2.28	10.49	3

gravity assist. However, its leverage factor is low, and it is interesting only for low-thrust missions such as BepiColombo (see Section 10.22.7).

With ΔVGA strategies, missions to Jupiter, comets, and Mercury would be within reach even if there was no other planet in the solar system. As we have seen in Section 10.22.2, Venus and Mars are relatively easy to reach. Tapping the gravity-assist potential of these two planets much improves the mass budget for more difficult targets, but there are additional constraints as now the Earth, Venus or Mars, and the target body must be in the proper relationship.

Venus is particularly attractive as a stepping stone to the inner solar system as this planet is nearly as large as the Earth. Furthermore, it lies nearly halfway to Mercury in terms of both its semimajor axis (0.723 AU, when Mercury is at 0.387 AU) and its orbit plane (inclination of 3.39°, to be compared with the 7° inclination of Mercury's orbit). A close swingby of Venus changes the velocity by as much as 7 km s^{-1}, which is more than what can be provided by a single rocket stage. All missions to Mercury therefore include at least one swingby of Venus, with departure velocities from the Earth in the 3.5–4 km s^{-1} range, much lower than the 7.5 km s^{-1}, which would be required for a direct transfer. Indeed, the Mariner 10 flyby mission to Mercury was the very first mission to use a planetary gravity assist, with a launch in November 1973, a flyby of Venus in February 1974, and three flybys of Mercury in 1974 and 1975 (Shirley, 2003).

For rendezvous missions, two Venus swingbys are required (Yen, 1989). The shortest solutions involve the 1:1 resonance (period of 0.615 years); the cheapest solutions use the 3:4 resonance (period of 0.461 years, coming back to Venus after 1.84 years). A second swingby then sends the spacecraft to Mercury, which is reached after 1.5–3.5 full orbits around the Sun (Mercury is usually not at the best position for a transfer, which requires inserting 'phasing' orbits). The arrival velocity is ~5.8 km s^{-1} in the best cases. This is still far too high for an orbiter mission: the corresponding capture maneuver would be in the 3 km s^{-1} range and could not be accommodated in the short time available close to this small planetary body (2450 km in radius). Fortunately, a sequence of gravity assists can also be used for reducing the arrival velocity, with the same leverage factors as for pumping up the departure velocity.

This strategy is particularly effective for Mercury, which has a short orbital period (0.24 years). Mercury mission strategies developed by Yen (1989) involve a series of three ΔVGA (2:1, 4:3, and 6:5) with a total cost of 500 m s^{-1} and a duration of 12 Mercury periods (slightly < 3 years), which reduce the relative velocity at arrival from 5.8 km s^{-1} down to 2.2 km s^{-1}, at which stage an insertion maneuver to a highly elliptical orbit requires only 500 m s^{-1}. Mercury has no atmosphere. Therefore, one has to rely on the spacecraft engine to further reduce the orbital period. The total ΔV cost of a Mercury orbiter on a 10 h elliptical orbit is now in a manageable range (1.5–2 km s^{-1}), which brings more than 50% of the launch mass into orbit, but cruise durations now range from 5.6 years to more than 9 years depending on the Venus window and the Venus–Mercury relative positions. This strategy is that used by the Messenger mission of NASA (McAdams et al., 2006; Santo et al., 2001), launched in 2005 and orbiting Mercury since 2011 (Figure 4).

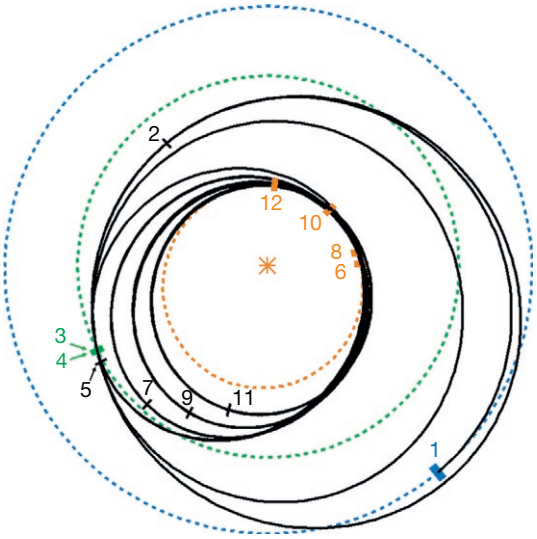


Figure 4 Messenger mission to Mercury, launched in August 2005, with an insertion in Mercury orbit in March 2011. After the launch at 4.05 km s^{-1} (1) and a deep space maneuver (2) that increases the relative velocity to Venus, two Venus swingbys are implemented (3,4). A small maneuver (5) is needed to reach Mercury (6). After the first flyby of Mercury (6) with a relative velocity of 5.8 km s^{-1}, a 3:2 ΔVGA (6–7–8), a 4:3 ΔVGA (8–9–10), and a 6:5 ΔVGA (10–11–12) are implemented. These three maneuvers lower the relative velocity to 2.2 km s^{-1} at a cost of only 500 m s^{-1} in ΔV (but more than 3 years in cruise time). The insertion maneuver into Messenger's elliptical orbit is now at an acceptable level (870 m s^{-1}).

Mars is not a very useful stepping stone to the outer solar system, as its mass is much smaller than that of Venus (the maximum ΔV available from a Mars gravity assist is ~3 km s^{-1}). Furthermore, the orbit of Mars is much closer to that of the Earth than to that of Jupiter. Conversely, Jupiter has such a large circular velocity (42 km s^{-1}) that it can sling the spacecraft toward any target in the outer solar system if the relative position is favorable. The Voyager 1 mission, launched in 1977, took advantage of a perfect configuration to visit Jupiter, Saturn, Uranus, and Neptune from 1980 to 1989, using each planet as a stepping stone to the next target. This favorable configuration repeats only once in 170 years. Given the relative periods of Jupiter (11.9 years) and Saturn (29.6 years), a favorable configuration between Jupiter and Saturn repeats every 19.8 years. The next visit to Saturn, the Cassini–Huygens mission, was launched in 1997, 20 years after Voyager, and the next programmatically realistic opportunity corresponds to launches in 2035–36. In the 18 years' intervals without Jupiter stepping stone opportunities, missions to Saturn must rely on gravity assists of inner planets combined with deep space maneuvers.

Indeed, the best use of the gravity potential of Venus and Mars for outer solar system missions is to replace the deep space maneuver by a gravity assist in an Earth–Earth ΔVGA. One can then reach large departure velocities from the Earth without expending any onboard ΔV, which much improves the mass budget (see, e.g., Longuski and Williams, 1991; Petropoulos et al., 2000).

A Mars flyby is most effective when providing the deep space maneuver in a 3:2 Earth ΔVGA scenario. The cheapest opportunity, with a departure velocity of ~3.2 km s^{-1}, occurs every 2.1 years, and the Earth return is 2.8 years later at a relative velocity of 9 km s^{-1} (Hechler, 1997). A 2 years' intermediate orbit can be needed so as to adjust the inclination. The cruise time is then increased by 5 years. A version of this strategy was used for the 2004 launch of Rosetta by Ariane 5. An initial 1:1 ΔVGA was added so as to optimize the launch declination while increasing the relative velocity from 3.4 to 3.85 km s^{-1} (**Figure 5**). With this strategy, the total ΔV is only 1.68 km s^{-1}, and the useful mass in rendezvous is 1400 kg, which makes it possible to implement a large orbiter and a lander on the comet nucleus. The drawback is a very long cruise time (10.5 years). This trade-off between mass budget and cruise time, already encountered for Mercury orbiter missions, is inherent to gravity-assist strategies: given enough time, very little is impossible.

A Venus gravity assist can provide for free a very large ΔV (up to 7 km s^{-1}). Therefore, a Venus flyby is already quite effective when used as the maneuver in a 1:1 Earth ΔVGA. Indeed, every 1.61 years, it is possible to depart to Venus with a velocity of 3.2–3.7 km s^{-1}, swing by Venus at a very safe distance, and then return to the Earth after 1–1.3 years with a relative velocity of 7.5–10 km s^{-1}, which is in the range required for reaching Jupiter and Saturn (see **Table 2**). The period of the return orbit is very close to 1 year. Hence, this large relative velocity is nearly perpendicular to the orbital velocity of the Earth, when the maximum rotation angle from an Earth flyby at this velocity as derived by eqn [11] is only 40°–50°. Therefore, a first Earth gravity assist is used to transfer the spacecraft to an orbit with a 2 years' period, during which a small maneuver can be implemented so as to further increase the relative velocity (2:1 ΔVGA). The second Earth gravity assist sets the spacecraft on its way to Jupiter or to the target comet. The cruise time is 3 years longer than for a direct transfer, but the low departure velocity provides large mass budgets (5 t approaching Jupiter with Ariane 5). This strategy was first implemented for sending the Galileo spacecraft into orbit around Jupiter (Damario et al., 1992), as it was much larger (1.6 t dry mass and 2.55 t at departure) than Pioneer or Voyager. Another example of such a scenario is the JUICE mission to the system of Jupiter launched in June 2022, which has been selected as the next large mission of the science program of ESA (**Figure 6**). Ariane 5 has strongly reduced launch capabilities away from the optimum declination (0°–5° South) as its third stage cannot be reignited. This increases the cruise time of JUICE by 1 year (up to 7.6 years) as one has to launch in June 2022 on a 1-year orbit at the optimum declination and then return to Earth 1 year later so as to rotate the relative velocity to the departure declination of 20.2° S required for reaching Venus in late November 2023. Europe approved in November 2012 the development of a restartable third stage for its new generation of launchers, which will much improve the capabilities of this launcher for planetary missions.

A different type of Venus–Earth gravity-assist strategy followed by a Jupiter swingby was used for the Cassini–Huygens mission to Saturn and Titan launched in 1997 (Peralta and Flanagan, 1995). A mission to Saturn ('Tandem') was considered as an alternative to JUICE in the selection process of ESA.

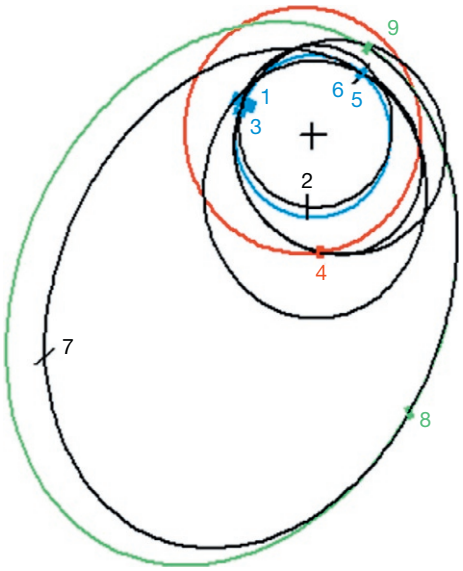


Figure 5 Trajectory of the Rosetta spacecraft to comet Churyumov–Gerasimenko (green orbit). After a launch by Ariane 5 in late February 2004 with a departure velocity of 3.4 km s^{-1} (1), a 1:1 ΔVGA maneuver (2) increased the relative velocity at the Earth to 3.85 km s^{-1} (3); a 3:2 Earth–Mars–Earth gravity assist (3–4–5) increased the velocity of the Earth encounter to 9.35 km s^{-1} (5); a sequence of 2 Earth flybys (5 and 6) is required to tilt the orbit so as to match the inclination of the comet. Two deep space maneuvers (7 and 8) totaling 1.45 km s^{-1} implement the rendezvous at 3.7 AU from the Earth in September 2014 (8). The comet is followed until perihelion (9) in October 2015 in the nominal mission.

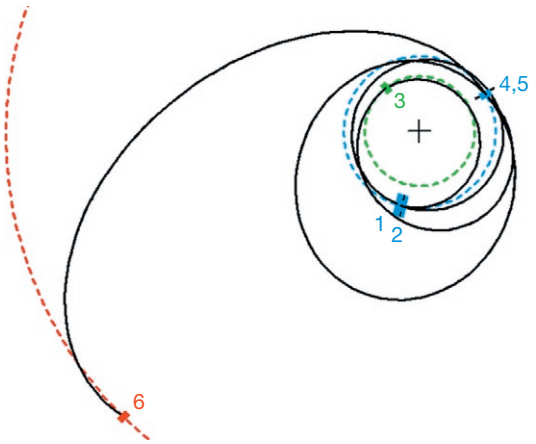


Figure 6 Nominal trajectory (Earth–Venus–Earth–Earth gravity assist) for JUICE, a mission to the Jupiter system selected in April 2012 as the large mission of the 'Cosmic vision' program of ESA. The launch at 3.32 km s^{-1} in June 2022 (1) is performed at the optimum declination for Ariane 5. A first swingby of the Earth 1 year later (2) sends the spacecraft to Venus. The Venus swingby (3) increases the relative velocity to 9.41 km s^{-1} when returning to the Earth (4). A third Earth flyby after 2 years (5) sends the spacecraft to Jupiter, which is reached in February 2030 after 7.6 years (6). Reproduced from Boutonnet A and Schoenmakers J (2012) Mission analysis for the JUICE Mission. *Advances in the Astronautical Sciences* 143: 1561–1578.

For launches in the 2020–25 time frame, Jupiter is too far ahead on its orbit to be of any use, and 'direct' Earth–Venus–Earth–Earth ΔVGA strategies to Saturn had to be considered. A small ΔVGA maneuver on a 4 years' orbit following the 2 years' orbit makes it possible to reach Uranus (Arridge et al., 2012) with a large spacecraft without relying on a Jupiter gravity assist, but the actual departure from the Earth is then 7.6 years after launch and the cruise time becomes extremely long (~15 years).

Gravity assists provide very effective means of exploring the environment of a planet if one or more large natural satellites are present. A remarkable example is the 4 years' tour of the Saturn system by Cassini (Wolf and Smith, 1995), with its 50 flybys of Titan, the largest satellite of Saturn. All other major satellites are visited, the rings are observed at close range, and high latitudes in the magnetosphere are explored. This tour has been extended until 2017, with more than 50 additional flybys of Titan and other satellites. The Jupiter system provides outstanding opportunities (Longman and Schneide, 1970) with its four Galilean satellites (Io, Europa, Ganymede, and Callisto), all with a large gravity-assist potential as escape velocities are larger than 2 km s^{-1} (**Table 6**); hence, velocity changes of 1 km s^{-1} or more can be achieved at no cost other than navigation corrections. This was already demonstrated by the Galileo mission of NASA (Damario et al., 1992). A complex sequence of ΔVGA and gravity-assist strategies between Galilean satellites can be implemented in relatively short times, as the period of Ganymede (third from Jupiter) is only 7.16 days instead of 365.25 days for the Earth (third from the Sun). Io, Europa, and Ganymede are tied by the Laplace resonance (periods close to 1:2:4), while the orbital period of Callisto is 16.7 days, 2.35 times that of Ganymede. The issue of windows from Ganymede to Callisto and back (an essential element of the JUICE strategy, as demonstrated by **Figure 6**) was addressed by Langevin (2010). An important constraint for designing the JUICE strategy (Boutonnet and Schoenmakers, 2012) was the integrated radiation dose, which increases very fast when the trajectory gets significantly closer to Jupiter than the orbit of Ganymede near the equatorial plane of the system.

An effective scenario has been designed (Boutonnet and Schoenmakers, 2012) for inserting the JUICE spacecraft in orbit around Ganymede after visiting Europa and then Callisto (Io flybys could not be considered due to the impact on radiation dose). This 2.6 years' tour (**Figure 7**) implements 11 Ganymede flybys, 13 Callisto flybys, and 2 Europa flybys before insertion in an elliptical orbit around Ganymede with an acceptable radiation budget. The available dry mass is 1800 kg (including 20% of system margins) with an Ariane 5 launch.

10.22.7 Advanced Propulsion Systems: Solar Sail and Ion Propulsions

Gravity-assist techniques presented in **Section 10.22.6** have made possible a wide range of missions, which would have been out of reach of a direct launch strategy. It is therefore interesting to present the arguments for new propulsion

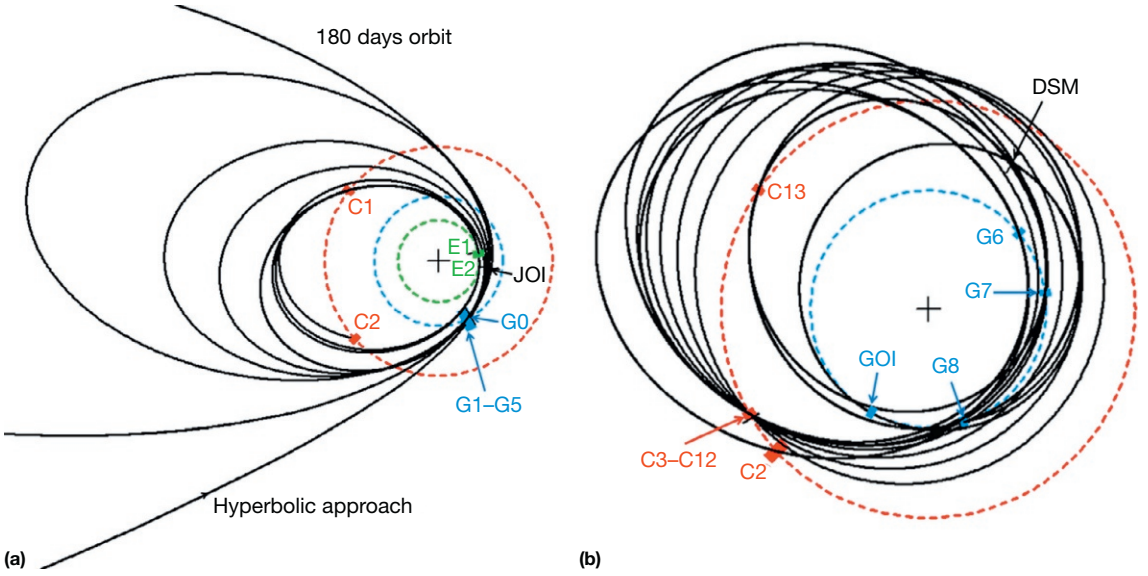


Figure 7 (a) Early stages of the Jupiter satellite tour of JUICE: a Ganymede swingby (G0) before Jupiter orbit insertion (JOI) reduces the JOI ΔV from 1.2 to 0.9 km s^{-1} for reaching a 180-day orbit. 5 Ganymede swingbys (G1–G5) and a small apocenter maneuver reduce the period of the orbit to $<$20 days. A sequence of two flybys of Europa (E1 and E2) bracketed by two flybys of Callisto (C1 and C2) can then be implemented. (b) Late stages of the tour: A Ganymede gravity assist (G6) reduces the relative velocity at Callisto from 4.9 km s^{-1} (C2) to 4.4 km s^{-1} (C3); a series of 10 polar swingbys (5 over the South pole and then 5 over the North pole) crank the inclination of a 16.7 days orbit (1:1 resonance with Callisto) up to 30° and then back to the orbital plane of the satellites; gravity assists by Ganymede (G7) and then Callisto (C13) reduce the relative velocity to $<$1.7 km s^{-1} at return to Ganymede (G8). A Ganymede–Ganymede ΔVGA approach ('DSM') on a 4:3 resonant orbit (G8 → GOI) is presented, which reduces the ΔV cost of Ganymede Orbit Insertion (GOI). A more complex strategy with two additional Ganymede flybys saves even more fuel. Reproduced from Boutonnet A and Schoenmakers J (2012) Mission analysis for the JUICE Mission. *Advances in the Astronautical Sciences* 143: 1561–1578.

techniques. It should first be noted that several missions of interest remain nearly impossible or require very long cruise times:

- Rendezvous with a small body (asteroid or comet) if the perihelion is not close to 1 AU or if the inclination exceeds 15° (the rendezvous maneuver exceeds the 3 km s^{-1} limit); this is the case for all main belt asteroids and all but a few comets.
- Rendezvous missions to Uranus, Neptune, and KBOs (including Pluto).

An additional problem is the exceedingly long cruise time, with an impact on the risk/reliability trade-off (hence cost). The relevance of missions designed 20 years before the science return is obtained is also an issue. Such missions cannot benefit from the latest advances in instrument design. They cannot adjust to new discoveries from other solar system exploration missions. Furthermore, there is an important programmatic risk, as other space agencies may try to steal the fire out of the mission with a smaller, faster mission. As an example, NASA programmed and then canceled a comet lander mission, which would have arrived 5 years earlier than the lander of Rosetta. The complexity of the Rosetta mission strategy (**Figure 6**) and that of the JUICE mission strategy (**Figure 7**) demonstrates that gravity-assist techniques have reached their limit with classical propulsion systems.

The only solution for improving this situation is a major increase in onboard ΔV capability, which requires increasing the ejection velocity V_{ej}. There is, however, a catch. The thrust T is

$$T = \mathrm{d}m/\mathrm{d}t \cdot V_{ej} \quad [20]$$

where $\mathrm{d}m/\mathrm{d}t$ is the mass flux in kg s^{-1}. The corresponding power W is

$$W = 0.5\,\mathrm{d}m/\mathrm{d}t \cdot V_{ej}^2 \quad [21]$$

hence,

$$W = 0.5 V_{ej} T \quad [22]$$

For a given thrust, the required power is therefore proportional to the ejection velocity. Large ejection velocities are appealing as they reduce the fuel mass for a given ΔV, but one must be aware of the larger corresponding energy requirements.

10.22.7.1 Solar Sail Propulsion

The 'photon drive' (e.g., a laser) is popular in science fiction novels, and it provides the highest possible ejection velocity (c, the speed of light). At relativistic speeds, the relationship between energy and thrust is less favorable than that given in eqn [22], as we have

$$W = cT \quad [23]$$

From eqn [23], an effective power of 300 MW would be required for a modest 1 N thrust (420 MW onboard power if the conversion efficiency is ~70%). Unless one considers direct mass–energy conversion (yet another science fiction favorite), the only realistic propulsion method using photons consists in implementing a solar sail, which reflects solar photons. This looks very appealing, in particular for very demanding missions such as sample returns (Hughes et al., 2006), as the photons are provided for free by the Sun; hence, there is no mass expenditure whatever the total ΔV. However, the available power at 1 AU is 1400 W m^{-2}; hence, a 200 000 m^2 sail (nearly 500 m × 500 m) is required for a 1 N thrust level. A maximum of 2 g m^{-2} for the sail and structure is required so as to keep the mass of the sail below 1000 kg. Furthermore, reflective properties must be maintained against high radiation levels for the full cruise phase (Edwards et al., 2004). Due to these difficulties, no solar sail mission has yet been flown. Several test missions have been considered by NASA and ESA, but the first test of solar sail propulsion is yet to be approved.

10.22.7.2 Ion Propulsion

Accelerating ions instead of molecules (ion propulsion) provides ejection velocities in the 15–100 km s^{-1} range, which are much higher than that provided by chemical thrusters (3.1 km s^{-1} for onboard systems). The best propellant is a heavy noble gas, such as Xenon. There are two methods for accelerating ions: the expansion of a hot plasma generates ejection velocities in the lower part of the range (15–20 km s^{-1}). With a high-voltage grid, it is possible to accelerate individual ions to velocities in the 30–100 km s^{-1} range. Higher ejection velocities require less propellant for a given ΔV. However, the required installed power increases with ejection velocity, which has an impact on the mass of the power subsystem. Electric energy can be obtained in space either from solar panels (solar electric propulsion, SEP) or from a small nuclear reactor (nuclear electric propulsion, NEP).

At 1 AU, a solar panel provides ~280 W m^{-2} with a 20% efficiency, which is already quite good. This corresponds to 26 mN with a 15 km s^{-1} ejection velocity and 8 mN with a 50 km s^{-1} ejection velocity, assuming an efficiency of 70% for the ion acceleration process. A SEP propulsion system is therefore limited to relatively small thrusts (in the range of 100–200 mN at 1 AU); otherwise, the solar panels become very large, which has an impact in terms of mass and cost. A 200 mN thrust seems very small, but if it is sustained over 1 year, it can provide 3.15 km s^{-1} of ΔV to a 2000 kg spacecraft. SEP missions are therefore characterized by very long thrust arcs.

SEP systems have been successfully tested by NASA (DS-1, which encountered comet Borelly), ESA (Smart-1, in orbit around the Moon from early 2005 to 2006; Racca et al., 2002), and Japan (Hayabusa, which performed a rendezvous with asteroid Itokawa in 2005 then returned to Earth in 2010). Two major missions have already been programmed: Dawn, launched in 2007 by NASA, has successfully reached 4 Vesta in July 2011 and has departed in September 2012 to Ceres after 1 year of observations of Vesta from orbit. BepiColombo (ESA/Japan) is a dual-orbiter mission to Mercury, which will be launched in 2016 (Benkhoff et al., 2010).

Even if onboard ΔV is more plentiful than with chemical missions, gravity-assist strategies can also be quite interesting for SEP missions, as demonstrated by the evolution of the mission scenario for BepiColombo, which initially included two orbiters and a lander. The very first studies considered a

direct spiral from the orbit of the Earth down to the orbit of Mercury of a large composite spacecraft with all three science modules. Even with an Ariane 5 launch, this required more than 16 km s^{-1} of ΔV with relatively high thrust levels (400 mN) and a 2.5 years' cruise time. These requirements exceeded the lifetime of thrusters, so that full spares were needed with a drastic impact on the cost and mass budgets, and this option was abandoned. Inserting a double Venus swingby reduced the ΔV by ~7 km s^{-1}, and a final 180° transfer from Mercury to Mercury, followed by a 1:1 ΔVGA, saved an additional 1.5 km s^{-1} (Langevin, 2000), so that the ΔV budget went down to a manageable level of ~7.5 km s^{-1} (Figure 8).

The mass requirements of subsystems rose much faster than usual during the early phases of the study (in particular for a Mercury lander; see Section 10.22.5), and cost became an issue. Two Soyuz–Fregat launches were considered, one for the lander and one for the two orbiters. The largest mass is then obtained by launching on a Moon-crossing orbit (~2200 kg).

The cost restrictions on cornerstone missions led in 2003 to the cancelation of the lander, which required a full Soyuz launch with a dedicated cruise module and ion propulsion module (see Section 10.22.5). The new configuration of the mission consisted of two orbiters launched by a single Soyuz–Fregat. Unfortunately, subsystem mass (in particular that of the SEP stage) continued to grow, which led to a major mass crisis, and ESA had to switch to an Ariane 5 launch (4100 kg launch capability, to be compared with 2200 kg for Soyuz–Fregat). New scenarios had to be considered so as to further reduce ΔV (hence fuel mass) with a thrust to mass ratio much less favorable than for early mission profiles. The present class of trajectories (Figure 9) looks quite similar to a ballistic mission such as Messenger (Figure 4) as more work is done by gravity assists (Earth, Venus, and Mercury) than by SEP thrust. Actually, 35% of the ΔV is spent accelerating the spacecraft (green arcs) instead of braking it (red arcs) contrary to the spiral-like approach initially considered (Figure 8). Thanks to the low fuel consumption of electric propulsion at the selected ejection velocity (45 km s^{-1}), more than 3400 kg of the 4000 kg launched in a Moon-crossing orbit can be delivered at a very low relative velocity to Mercury, but the cruise time ranges now from 6.4 to 7.8 years (it was only 2.6 years with Ariane 5 when the mass budget was much more favorable).

The evolution of BepiColombo illustrates the remarkable flexibility of electric propulsion: contrary to chemical propulsion missions, the large onboard ΔV capability makes it possible to adjust to major changes in mass budgets or launch dates. The combination of gravity assists with low thrust is extremely effective, as the initial ΔV budget of 16 km s^{-1} could be reduced to ~4.5 km s^{-1}, albeit with a significant increase in cruise time (from 2.6 to 6.4 years). Additional information on SEP optimization methods can be obtained from Chesta et al. (2006), while the navigation aspects of BepiColombo are addressed by Yarnoz et al. (2006) and Alessi et al. (2012).

SEP is only effective relatively close to the Sun. The Dawn mission, which implements a rendezvous with Vesta, at

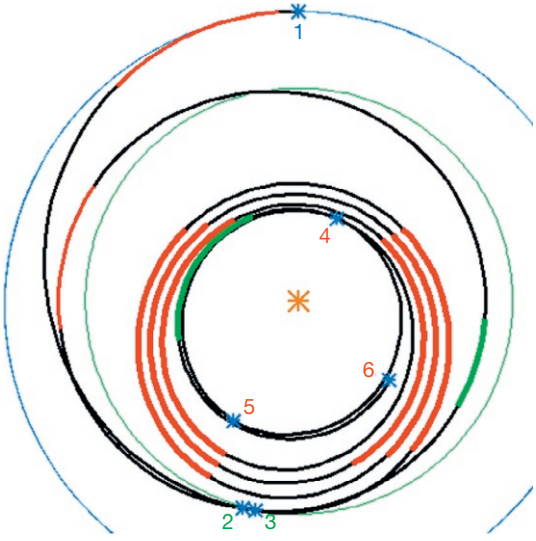


Figure 8 Direct launch option by an Ariane 5 for the BepiColombo mission to Mercury in late 2009. The departure velocity of 2.5 km s^{-1} is increased by early thrusting with the ion engine, so as to reach Venus with a relative velocity of 4.5 km s^{-1} (2). A 1:1 Venus ΔVGA increases the relative velocity to 5.7 km s^{-1} (3). After a three-orbit spiral, Mercury is reached at a relative velocity of 2.7 km s^{-1} (4). A 180° transfer (4–5) and then a 1:1 Mercury ΔVGA (5–6) reduce the relative velocity to $<$400 m s^{-1}. Red arcs correspond to thrust against the orbital velocity; green arcs correspond to thrust along the orbital velocity. The cruise time is 2.6 years. Reproduced from Langevin Y (2000) Chemical and solar electric propulsion options for a cornerstone mission to Mercury. *Acta Astronautica* 47: 443–452.

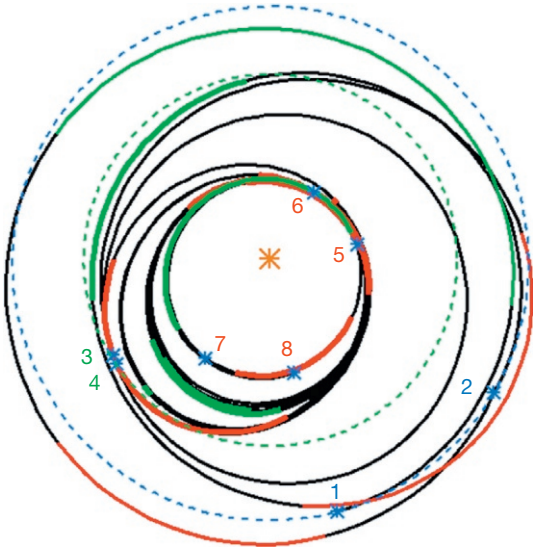


Figure 9 Example of a class of low ΔV SEP transfers to Mercury with an Ariane 5 launch, which constitutes the new baseline for BepiColombo (now scheduled for a launch in 2016). The two thrust arcs on the 1:1 orbit from departure (1) to the Earth swingby (2) effectively increase the relative velocity (1:1 ΔVGA). The Earth swingby provides the large declination (37° S) required for reaching Venus on a type 3 transfer. The two Venus swingbys combined with some braking thrust (red arcs) make it possible to reach Mercury (5). A ΔVGA (green arcs) on a 5:4 orbit (5–6) combined with braking arcs near perihelion reduces the relative velocity to Mercury down to 1.8 km s^{-1}. A 180° transfer (6–7) further reduces it to 1.2 km s^{-1}, at which stage an end-game strategy using 3-body effects (Jehn et al., 2008) cheaply inserts the BepiColombo composite in orbit around Mercury (8).

2.4 AU, and then Ceres, at 2.8 AU, already requires very large solar panels. Therefore, SEP is totally inappropriate for outer solar system missions. Nuclear reactors are the only power source that can provide several 10 kW at large distances from the Sun, so as to power ion thrusters (NEP). The smallest nuclear reactors provide several 100 kW, with a mass of several tons. The scale of NEP missions is daunting: JIMO, the mission to Galilean satellites, which was extensively studied in 2000–05, consisted of an 8 t spacecraft, with 5 t for the reactor and shield, and a thrust of 8 N. In the outer solar system, a JIMO level acceleration (1 mm s^{-2}) far exceeds the gravitational pull of the Sun (0.23 mm s^{-2} at 5 AU). Therefore, trajectories are non-Keplerian, which makes it possible to reach Saturn, Uranus, or Neptune in a fraction of the time, which would be required on a Hohmann transfer orbit (8 years for Saturn and 29 years for Neptune). Such fast trajectories require ΔV budgets >20 km s^{-1}, as one has to accelerate and then brake at arrival. Fortunately, ejection velocities can be very high (~100 km s^{-1}) as there is a lot of available power; hence, 20 km s^{-1} of ΔV only uses up ~20% of the mass as fuel. The cancelation of JIMO by NASA in 2005 was a severe setback for NEP, which is undoubtedly an enabling technology for a systematic exploration of the outer solar system beyond Saturn.

10.22.8 The Specific Challenges of Sample Return Missions

Implementing a sample return mission is a very challenging task, as one can consider it as a rendezvous mission followed by its mirror image. This view is to a large extent correct. However, there are some mitigating circumstances. First of all, the Earth has a dense atmosphere, which is quite capable of braking sample return modules at large velocities. It is 'only' required to set the spacecraft (or the sample return capsule, provided that it can survive alone) on a trajectory that intersects that of the Earth. In terms of ΔV budget, one gets rid of the huge cost of leaving the sphere of influence of the Earth. However, one has to launch from the target body.

Asteroids and comets represent relatively easy targets for a landing and sample collection, as escape velocities are either low or negligible. A specific case is constituted by the Stardust mission, which brought back samples from comet Wild-2 in early 2006. After a launch in early 1999, a 2 years' ΔVGA maneuver increased the relative velocity so that an Earth gravity assist in January 2001 set Stardust on its way to Wild-2 on an orbit with a period of nearly 2.5 years. The samples were collected at a high velocity (6.1 km s^{-1}) during a flyby in early 2004, and a small maneuver was then required to encounter the Earth in January 2006, 5 years after the Earth gravity assist, after two full orbits. Due to the high impact velocity, only very small grains can be captured relatively unaltered. For a comet nucleus sample return (once the baseline of Rosetta that was then considered in collaboration with NASA), two large maneuvers are needed so as to rendezvous with the comet and then leave it toward the Earth. Figure 5 illustrates the specific difficulty of implementing a comet nucleus sample return with a chemical propulsion mission: cheap rendezvous strategies arrive after aphelion, and by symmetry, cheap returns depart before aphelion. One would need to wait nearly a full orbit around the comet. Hence, the duration would be that of a rendezvous mission plus 6 years (typical period for a comet) plus 3 years for the return, for a total of 15–20 years. Much shorter sample return missions can be achieved with ion propulsion, as the rendezvous then occurs after perihelion; hence, the return trip can be initiated almost immediately (e.g., Kawaguchi, 2002). An example is given in Figure 10 for comet Wild-2. It constitutes a first illustration of the mirror-image character of the return trip. The 1:1 ΔVGA loop after launch can be skipped for the return leg as there is no need to reduce the relative velocity to the Earth, which will be cheaply dissipated in the atmosphere of our planet.

The very first SEP sample return mission has been the Japanese Hayabusa missions. After a launch in 2003, it achieved its rendezvous with asteroid Itokawa in September 2005. Contact with the spacecraft was lost and then recovered after many attempts. SEP again proved its versatility as a new return strategy could be found, and a few mg of asteroid dust could be recovered in 2010. As Itokawa is an Earth-crossing asteroid, the ΔV budget was relatively low. The Japanese space agency (JAXA) has scheduled a reflight (Hayabusa 2) to a primitive Earth-crossing asteroid (162173 1999JU3), which will be launched in late 2014. A sample return from a main belt asteroid would have similar requirements ($\Delta V \sim 10$ km s^{-1}) and the same mirror spiral trajectory as a sample return from a cometary nucleus.

Among large bodies, the Moon is by far the easiest target for a sample return mission. It has a relatively low escape velocity and surface gravity (see Table 6), which more than offset the lack of an atmosphere. Furthermore, the trip to the Moon and back is very cheap: a Moon-crossing orbit is obviously adequate. The approach (or departure) velocity is 0.8 km s^{-1}, so that the ΔV for reaching or leaving a low circular lunar orbit from a transfer orbit is only 0.8 km s^{-1} (the two are equal by coincidence). Table 8 shows that the ratio between the mass approaching the Moon and the landed mass is in the 3.5–4

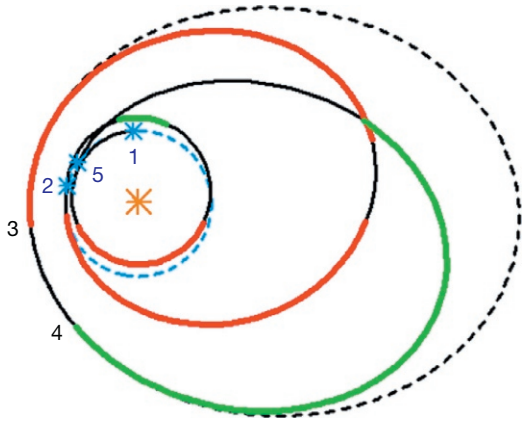


Figure 10 SEP comet sample return mission to Wild-2. A lunar swingby in December 2022 provides a departure velocity of 1.4 km s^{-1} (1). A two-arc 1:1 ΔVGA increases the relative velocity to 5.9 km s^{-1} (2). Rendezvous (3) is achieved in September 2021 by accelerating nearly continuously for 2.5 years (red arcs). After 3 months near the comet, the SEP propulsion system is reactivated, braking during 3 years (long green arc) so as to deliver the return capture in February 2031. The total ΔV is an acceptable 10.14 km s^{-1}; the round-trip duration is 9.14 years.

range, with a similar ratio between the mass leaving the lunar surface and the mass escaping toward the Earth. Therefore, a 50 kg return capsule requires only about 800 kg approaching the Moon, with one rocket stage for landing and one rocket stage for departure (similarly to the Apollo Landing Module). Indeed, the soviet missions Luna 16, Luna 20, and Luna 24 of the 1970s are until now the only successful automatic sample return missions from a large body (including a 2 m deep core), at a small fraction of the cost of the Apollo program.

The next easiest targets for a sample return mission are expected to be the two planets that are easiest to reach for a rendezvous, Mars and Venus. In both cases, it is possible to use aerobraking at arrival, which improves the mass budget. However, Mars and even more so Venus are large planets, with escape velocities of 5.02 km s^{-1} and 10.36 km s^{-1}, respectively.

The escape velocity of Mars is even larger than that of Mercury (4.26 km s^{-1}). As we have seen in **Section 10.22.5**, such levels of ΔV are almost out of reach of a single stage. Trade-offs for such a mission are presented in Donahue et al. (2006). The circular velocity (3.553 km s^{-1} at the surface) is more manageable, as it is closer to the exhaust velocity of bipropellant propulsion systems. The logical scenario for a Mars sample return mission is therefore that that was considered for a joint mission between NASA and several European countries (France, Germany, Finland, etc.) before its cancelation in 2002:

- A large orbiter spacecraft is launched toward Mars, where it implements an aerocapture maneuver into a low circular orbit so as to maximize the mass in orbit. It carries the return capsule.
- A large direct entry module is launched independently. It carries a small surface to orbit module. Once on the surface, samples are collected and then transferred to the surface to entry module, which sends a small sample cannister in orbit.
- The orbiter collects the sample cannister, which is transferred to the sample return module.
- When the return window opens, the sample return module is launched toward the Earth.
- The sample cannister is recovered after a direct entry into the Earth atmosphere.

Figure 11 provides an example of a Mars sample return mission launched in late 2013. The symmetry of the mission design is immediately apparent. The ΔV budget for the return is in this case very low (2.03 km s^{-1} from a circular orbit to a departure velocity of 2.5 km s^{-1} toward the Earth). A mass budget of ~200 kg is probably achievable for the sample return module. Arrival at Mars (2) is during southern summer storm season, which is not a problem for the orbiter, but may constitute a risk for the lander. A launch in September 2013 with one full additional orbit around the Sun arrives earlier in the seasonal cycle, but the Mars operational sequence would need to be completed in 2 months so as to catch the end of the return window in early 2016, and even then the departure velocity would be 3.5 km s^{-1}, with a higher ΔV 2.5 km s^{-1} (plus gravity losses) for the sample return module.

Venus is a relatively easy case for the round-trip issue. **Figure 12** gives an example of such a mission launched in late 2016, which takes nearly 2 years. The problem is that Venus is almost as large as the Earth (escape velocity of 10.35 km s^{-1} instead of 11.78 km s^{-1}). The hot and very thick atmosphere (90 bars at the surface, 450 °C) adds to the difficulty. At arrival, one can use aerocapture so as to retain a large fraction of the mass budget in low circular orbit or on the ground. At departure, the sample return module will need a ΔV of 3.36 km s^{-1} (plus gravity losses) from a low circular orbit, which requires by itself a large propulsion system. Recovering anything from the surface is a major issue: This is energetically similar to a launch in Earth orbit, with mass ratios higher than

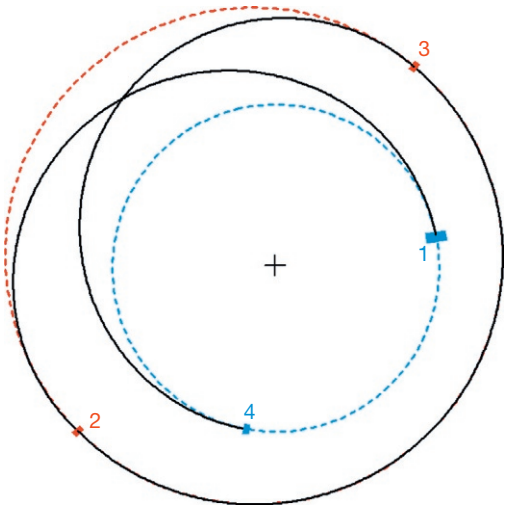


Figure 11 Favorable window for a Mars sample return mission, with a launch (1) in October 2024 (departure, 3.35 km s^{-1}, and Mars approach, 2.46 km s^{-1}). A type II transfer reaches Mars in September 2025 (2). One must wait 0.9 years at Mars (2–3), so as to reach a favorable type II return window, departing in August 2026 (3) at 2.68 km s^{-1} and delivering the sample in June 2027 (4), with a low Earth approach velocity (3 km s^{-1}). The nearly symmetrical round-trip takes 2.7 years.

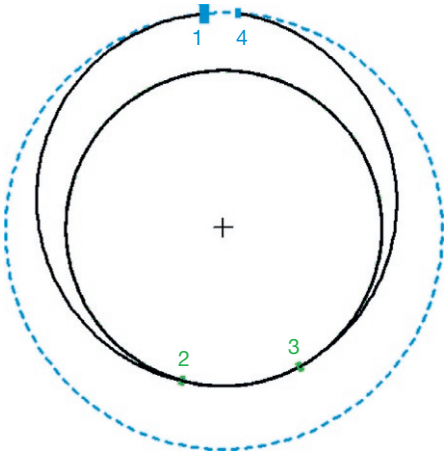


Figure 12 Venus sample return mission launched in January 2025 (departure velocity: 2.8 km s^{-1}). A type I transfer reaches Venus in April 2025. One must wait more than two full periods of Venus (1.2 years from 2 to 3) so as to reach a favorable type I return window, with a departure in August 2027 at 2.9 km s^{-1} and a sample delivery in December 2027, with a low approach velocity (4.25 km s^{-1}). The round-trip takes slightly <2 years. The symmetry between the departure and return legs is readily apparent.

20, but the launcher has to start under the equivalent of 900 m of water (400 m at the top of Maxwell Montes, the highest mountain of Venus). Until now, the record of survival at the surface, held by a Soviet Venera lander, stands at 2 h. It is therefore clear that enabling high-temperature technologies need to be validated by successful long-duration in situ missions (see, e.g., Landis, 2006) before addressing the daunting mass budget challenges of a Venus sample return mission from the surface itself. Recovering a sample from the upper atmosphere is a much less challenging task (Sweetser et al., 2003).

When considering the major technical issues, which had to be solved for an ambitious orbiter mission to Mercury and the very large mass budget required to land any useful mass on the surface, a sample return from Mercury appears very far-fetched indeed, unless solar sail technology becomes mature (McInnes, 2003). The same applies to large natural satellites of giant planets, where major steps in terms of orbiter mission and in situ studies need to be achieved before a sample return can be considered. Therefore, the most likely targets for sample return missions in the foreseeable future are small bodies (asteroids and comets), the Moon, and Mars, even if one cannot exclude a return of a sample of the upper atmosphere of Venus.

References

Albee AL, Arvidson RE, Palluconi F, and Thorpe P (2001) Overview of the Mars Global Surveyor mission. *Journal of Geophysical Research: Planets* 106: 23291–23316.

Alessi EM, Cicalo S, Milani A, and Tommei G (2012) Desaturation manoeuvres and precise orbit determination for the BepiColombo mission. *Monthly Notices of the Royal Astronomical Society* 423: 2270–2278.

Arridge CS, Agnor CB, André N, et al. (2012) Uranus Pathfinder: Exploring the origins and evolution of Ice Giant planets. *Experimental Astronomy* 33: 753–791.

Bate RR, Mueller DD, and White JE (1971) *Fundamentals of Astrodynamics*. New York: Dover Publications, ISBN: 0-486-60061-0.

Battin RH (1999) *An Introduction to the Mathematics and Methods of Astrodynamics*. AIAA (revised edition), ISBN: 1-56347-342-9: revised edition.

Beckers JM (1993) Adaptive optics for astronomy – Principles, performance and applications. *Annual Review of Astronomy and Astrophysics* 31: 13–62.

Belbruno EA and Miller J (1993) Sun-perturbed Earth-to-Moon transfers with ballistic capture. *Journal of Guidance, Control, and Dynamics* 16: 770–775.

Benkhoff J, van Casteren J, Hayakawa H, et al. (2010) BepiColombo – Comprehensive exploration of Mercury: Mission overview and science goals. *Planetary and Space Science* 58: 2–20.

Boutonnet A and Schoenmakers J (2012) Mission Analysis for the JUICE Mission. *Advances in the Astronautical Sciences* 143: 1561–1578.

Casalino L, Colasurdo G, and Pastrone D (1998) Optimization of Delta V Earth gravity-assist trajectories. *Journal of Guidance, Control, and Dynamics* 21: 991–995.

Cheng AF, Santo AG, Heeres KJ, et al. (1997) Near-Earth Asteroid Rendezvous: Mission overview. *Journal of Geophysical Research* 102: 23695–23708.

Chesta E, Estublier D, Fallis B, et al. (2006) Flexible variable-specific-impulse electric propulsion systems for planetary missions. *Acta Astronautica* 59: 931–945.

Craig S and Lyne JE (2005) Parametric study of aerocapture for missions to Venus. *Journal of Spacecraft and Rockets* 42: 1035–1038.

Damario LA, Bright LE, and Wolf AA (1992) Galileo trajectory design. *Space Science Reviews* 60: 23–78.

Donahue BB, Green SE, Coverstone VL, and Woo B (2006) Chemical and solar-electric propulsion system analyses for Mars sample return missions. *Journal of Spacecraft and Rockets* 43: 170–177.

Edwards DL, Hubbs WS, Wertz GE, et al. (2004) Electron radiation effects on candidate solar sail material. *High Performance Polymers* 16: 277–288.

Farquhar RW (1991) Halo orbits and lunar swingby missions of the 1990s. *Acta Astronautica* 24: 227–234.

Farquhar RW (2001) The flight of ISEE-3/ICE: Origins, mission history and a legacy. *Journal of the Astronautical Sciences* 49: 23–73.

Farquhar RW and Dunham DW (1999) The indirect launch mode: A new launch technique for interplanetary missions. *Acta Astronautica* 45: 491–497.

Farquhar RW, Kawaguchi J, Russell C, Schwehm G, Veverka J, and Yeomans D (2002) Spacecraft exploration of asteroids: The 2001 perspective. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 367–376. Tucson, AZ: University Arizona Press, part III.

French JR and Cruz MI (1980) Aerobraking and aerocapture for planetary missions. *Astronautics and Aeronautics* 18: 48–56.

Hechler M (1997) ROSETTA mission design. *Advances in Space Research* 19: 127–136.

Hechler M, Langevin Y, and Yanez A (2005) Mars Express orbit design evolution. *Dynamical Systems* 20: 175–188.

Hughes GW, Macdonald M, Macdonald M, McInnes CR, Atzei A, and Falkner P (2006) Terrestrial planets sample return mission using solar sail propulsion. *Acta Astronautica* 59: 796–806.

Jehn R, Companys V, Corral C, Yarnoz GD, and Sanchez N (2008) Navigating BepiColombo during the weak-stability capture at Mercury. *Advances in Space Research* 42: 1364–1369.

Justus CG, Duvall A, Keller VW, Spilker TR, and Lockwood MK (2005) Connecting atmospheric science and atmospheric models for aerocapture at Titan and the outer planets. *Planetary and Space Science* 53: 601–605.

Kawaguchi J (2002) A comet nucleus sample return mission using solar electric Earth gravity-assist. *Advances in Space Research* 29: 1215–1220.

Knezevic Z and Milani A (1998) Orbit maintenance of a lunar polar orbiter. *Planetary and Space Science* 46: 1605–1611.

Landis GA (2006) Robotic exploration of the surface and atmosphere of Venus. *Acta Astronautica* 59: 570–579.

Langevin Y (2000) Chemical and solar electric propulsion options for a cornerstone mission to Mercury. *Acta Astronautica* 47: 443–452.

Langevin Y (2001) Solar system: Exploration. In: Murdin P (ed.) *Encyclopedia of Astronomy and Astrophysics*, pp. 2754–2762. Institute of Physics and Nature publishing group.

Langevin Y (2005) Design issues for space science missions. In: Martinez-Pillet V, Aparicio A, and Sanchez F (eds.) *Payload and Mission Definition in Space Sciences*, pp. 17–89. Cambridge University Press, Chapter 2.

Langevin Y (2010) Mission design issues for the European orbiter of Laplace/EJSM: Callisto fly-by sequences. *Advances in the Astronautical Sciences* 135: 967–986.

Longman RW and Schneide AM (1970) Use of Jupiter moons for gravity-assist. *Journal of Spacecraft and Rockets* 7: 570–578.

Longuski JM and Williams SN (1991) Automated design of gravity-assist trajectories to Mars and the outer planets. *Celestial Mechanics* 52: 207–220.

Marsden JE and Ross SD (2006) New methods in celestial dynamics and mission analysis. *Bulletin of the American Mathematical Society* 43: 43–73.

McAdams JV, Dunham DW, Farquhar RW, Taylor AH, and Williams BG (2006) Trajectory design and maneuver strategy for the Messenger mission. *Journal of Spacecraft and Rockets* 43: 1054–1064.

McInnes CR (2003) Low cost Mercury orbiter and sample return missions using solar sail propulsion. *Aeronautical Journal* 107: 469–478.

Peralta F and Flanagan S (1995) Cassini interplanetary trajectory design. *Control Engineering Practice* 3: 1603–1610.

Petropoulos AE, Longuski JM, and Bonfiglio EP (2000) Trajectories to Jupiter via gravity assists from Venus, Earth and Mars. *Journal of Spacecraft and Rockets* 37: 776–783.

Racca GD, Marini A, Stagnaro L, et al. (2002) SMART-1 mission description and development status. *Planetary and Space Science* 50: 1333–1337.

Santo AG, Gold RE, McNutt RL, et al. (2001) The Messenger mission to Mercury: Spacecraft and mission design. *Planetary and Space Science* 49: 1481–1500.

Scoon G, Whitcomb G, Eiden M, and Smith A (1991) Cassini-Huygens entry and descent technologies. *Space Technology* 11: 167–177.

Shirley DL (2003) The Mariner 10 mission to Venus and Mercury. *Acta Astronautica* 53: 375–385.

Smith JC and Bell JL (2005) 2001 Mars Odyssey aerobraking. *Journal of Spacecraft and Rockets* 42: 406–415.

Spilker TR (2005) Significant science at Titan and Neptune from aerocapture. *Planetary and Space Science* 53: 606–616.

Sweetser T, Peterson C, Nilsen E, and Gershman B (2003) Venus sample return missions – a range of science, a range of costs. *Acta Astronautica* 52: 165–172.

Tu KY, Munir MS, Mease KD, and Bayard DS (2000) Drag-based predictive tracking guidance for Mars precision landing. *Journal of Guidance, Control, and Dynamics* 23: 620–628.

Vallado DA (2001) *Fundamentals of Astrodynamics and Applications*, 2nd edn. Hawthorne: Microcosm/Kluwer, ISBN: 1-881-88312-4.
Vinh NX and Kuo Z-S (1997) Improved matched asymptotic solutions for deceleration control during atmospheric entry. *Acta Astronautica* 40: 1–11.
Willcockson WH (1999) Stardust sample return capsule design experience. *Journal of Spacecraft and Rockets* 36: 470–474.
Wolf AA and Smith JC (1995) Design of the Cassini tour trajectory in the saturnian system. *Control Engineering Practice* 3: 1611–1619.
Wu SF, Costa RR, Chu QP, Mulder JA, and Ortega G (2001) Nonlinear dynamic modeling and simulation of an atmospheric re-entry spacecraft. *Aerospace Science and Technology* 5: 365–381.
Yarnoz D, Jehn R, and Croon M (2006) Interplanetary navigation along the low-thrust trajectory of BepiColombo. *Acta Astronautica* 59: 284–293.
Yen CS (1989) Ballistic Mercury orbiter mission via Venus gravity-assist. *Journal of the Astronautical Sciences* 37: 417–432.

This page intentionally left blank

10.23 Instrumentation for Planetary Exploration Missions

P Falkner and R Schulz, European Space Agency, Noordwijk, The Netherlands

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by P Falkner, A Peacock and R Schulz, Volume 10, pp. 595–641, © 2007, Elsevier B.V.

10.23.1	**Introduction**	720
10.23.2	**Building on Past Missions**	721
10.23.3	**Exploration Strategies and Associated Techniques**	722
10.23.3.1	Exploration Strategies for Near-Future Perspectives	722
10.23.3.1.1	Exploration of the Jovian system	722
10.23.3.1.2	Preparation of landing on terrestrial planets	723
10.23.3.1.3	Sample and return from a low-gravity environment	723
10.23.3.2	Longer-Term Goals for Exploration	724
10.23.4	**Instruments**	724
10.23.4.1	Measurement Principles	724
10.23.4.1.1	Direct detection	724
10.23.4.1.2	Indirect detection	724
10.23.4.1.3	Remote sensing	725
10.23.4.1.4	*In situ* sensing	725
10.23.4.1.5	Passive detection	725
10.23.4.1.6	Active detection	725
10.23.4.2	Imager	726
10.23.4.2.1	Wide field imager	726
10.23.4.2.2	High-resolution camera	726
10.23.4.2.3	Stereo imager	727
10.23.4.2.4	Scanning platform	727
10.23.4.2.5	Panoramic camera	727
10.23.4.2.6	Microscopic imager	727
10.23.4.3	Spectrometer	727
10.23.4.3.1	Visual and IR mapping spectrometer	728
10.23.4.3.2	UV spectrometer	728
10.23.4.4	Radiometer and Polarimeter	728
10.23.4.4.1	Thermal IR radiometers	729
10.23.4.5	Nephelometer	729
10.23.4.6	x-Ray Spectrometer	729
10.23.4.6.1	x-Ray fluorescence	730
10.23.4.6.2	Particle-induced x-ray emission	730
10.23.4.6.3	Alpha particle x-ray spectrometer	730
10.23.4.7	Gamma-Ray Spectrometer	730
10.23.4.8	Neutron Spectrometer	731
10.23.4.9	Laser Altimeter	731
10.23.4.10	Radar	732
10.23.4.10.1	Radar altimeter	732
10.23.4.10.2	Ground-penetrating radar	732
10.23.4.10.3	Synthetic aperture radar	733
10.23.4.11	Radio Science	733
10.23.4.11.1	Celestial mechanics and the gravitational field	733
10.23.4.11.2	Radio propagation at occultation	733
10.23.4.12	Microwave Sounder	733
10.23.4.13	Particle Detector	734
10.23.4.13.1	Heavy ion counter	734
10.23.4.13.2	Energetic particle analyzer	734
10.23.4.13.3	Neutral particle analyzer	734
10.23.4.14	Mass Spectrometer	734
10.23.4.14.1	Magnetic focusing analyzer	734
10.23.4.14.2	Quadrupole mass analyzer	735
10.23.4.14.3	Ion trap	735

10.23.4.14.4 Wien filter 735
10.23.4.14.5 Gas chromatograph–mass spectrometer 736
10.23.4.14.6 TOF mass analyzer 736
10.23.4.14.7 Laser ablation mass spectrometer 736
10.23.4.14.8 Laser-induced breakdown spectroscopy 736
10.23.4.14.9 Raman spectrometer 736
10.23.4.15 Mössbauer Spectrometer 737
10.23.4.16 Dust Analyzer 737
10.23.4.16.1 Quartz crystal microbalance 737
10.23.4.17 Radiation Environment Monitor 737
10.23.4.18 Scanning Tunneling and Atomic Force Microscope 738
10.23.4.19 Scanning Electron Microscope 738
10.23.4.20 Plasma and Wave Instruments 738
10.23.4.20.1 Plasma wave analyzer 738
10.23.4.20.2 Plasma analyzer 739
10.23.4.20.3 Langmuir probe 739
10.23.4.20.4 Electron drift instrument 739
10.23.4.20.5 Active spacecraft potential control 739
10.23.4.21 Magnetometer 739
10.23.4.21.1 Fluxgate magnetometer 740
10.23.4.21.2 Vector helium magnetometer 740
10.23.4.21.3 Search coil 740
10.23.4.22 Accelerometer and Seismometer 740
10.23.4.22.1 Seismometer 741
10.23.4.23 Thermometer 741
10.23.4.24 Pressure Sensor 741
10.23.4.25 Microphones 742
10.23.4.26 Microgas Sensors 742
10.23.4.27 Surface Dating 742
10.23.5 Instrument Suites for Exploration 742
10.23.5.1 Planetary Size, Mass, Shape, and Interior 743
10.23.5.2 Surface and Subsurface 743
10.23.5.2.1 Surface station 744
10.23.5.2.2 Rover 745
10.23.5.2.3 Subsurface access 745
10.23.5.2.4 Mechanical tools 745
10.23.5.3 Atmospheres, Exospheres, and Ionospheres 745
10.23.5.3.1 Entry and descent probes 746
10.23.5.3.2 Aerobots 747
10.23.5.4 Magnetosphere 747
10.23.5.5 Moons 748
10.23.5.6 Rings 748
10.23.5.7 Asteroids and Comets 749
10.23.5.8 Exobiology 750
10.23.5.9 Sample Return 750
10.23.5.10 Engineering Aspects and Operations 751
10.23.5.10.1 Engineering constraints 751
10.23.5.10.2 Operational aspects 751
10.23.5.10.3 Verification and testing 752
10.23.6 Outlook and Timeline 752
References 752

10.23.1 Introduction

The bodies making up our solar system are usually divided into different groups (see **Chapter 10.01**). Often, we simply distinguish between inner and outer planets, moons, and small bodies, namely, comets and asteroids. However, if we wish to find ways to investigate these bodies through planetary exploration missions, we need to first identify those of their basic properties that are relevant for the definition of the mission. Different solar system objects need different study strategies and therefore different types of instrumentation. Specifically, we have to distinguish between the measurements needed to understand

- terrestrial planets and large moons with atmosphere;
- terrestrial planets and large moons without atmosphere;
- giant planets and their environment, for example, Jovian magnetosphere, Io torus, and Saturn rings;
- dwarf planets and trans-Neptunian objects (TNOs);
- small bodies without atmosphere, for example, asteroids and small moons;
- comets; and
- interstellar and interplanetary dust particles.

Apart from the obvious need to design an exploration mission to any of these objects in accordance with the conditions expected during the transfer from Earth to the object and at the object's location (e.g., thermal and radiation environment), the mission scenario and scientific instrumentation need to be specifically tailored to the properties of the object itself. For instance, the requirements are very different for landing on a terrestrial planet or a small body with low gravity. Similarly, the requirements on exploring an airless body are very different from those required for a body with a thick atmosphere. In addition, the selection of the mission scenario depends decisively on what we already know of the object(s) from other investigations and past missions. For example, if a certain object has been visited and studied by remote-sensing techniques from orbit to an extent that no epoch-making results can be expected from another remote-sensing mission, the logical next step should be to obtain *in situ* measurements. Depending on the type of body, the *in situ* measurements would involve the study of the atmosphere, surface/subsurface, or both.

Generally, one can distinguish between three broad classes of measurements, namely, remote sensing, environmental sensing, and *in situ* sensing. For missions to solar system bodies (planets, moons, asteroids, and comets), remote sensing, even from Earth, would usually be the first step of any investigation. Remote sensing from a spacecraft largely benefits from the simple fact that the measurements can be obtained very close to the object to be investigated, rather than from observatories on Earth or in Earth orbit. Typical remote-sensing instruments would cover the classical measurement triangle: spatial (imager)/spectral (spectrometer)/intensity (radiometer). In addition, any mission focusing on remote sensing of a solar system body from orbit or during a flyby would also include a number of instruments for environmental sensing of any existing plasma or magnetosphere around the object. For the next step, *in situ* measurements, the choice of mission scenario, and payload depends critically on the object to be investigated, as well as on which aspect of this target shall be studied. For example, if we wish to perform *in situ* measurements of the atmosphere of a body, the requirements for the mission will be very different from those for *in situ* investigations of the (sub)-surface composition. For the latter, it is mandatory to actually land on the object under investigation, while the former might for some objects (comets, asteroids, and planets/moons with exosphere) be accomplished by simply adding to the payload instruments with *in situ* measurement capabilities, such as mass spectrometers. In other cases, particularly for bodies with dense atmospheres, such as Venus, Titan, and the gas giants, an atmospheric entry probe is required. An overview is provided on possible instrumentation for various types of planetary exploration missions adjusted to the strategies required for the respective investigations.

10.23.2 Building on Past Missions

Spacecraft have visited each of the eight planets of our solar system, as well as a number of moons, comets, and asteroids. The levels at which certain objects have been investigated go hand in hand with the types of missions that have been conducted so far. At present, samples from the Moon and cometary dust have been returned to Earth. The Stardust mission has collected dust particles from the coma of comet 81P/Wild 2. The sample return capsule with the dust particles stowed into aerogel was safely returned to Earth on 15 January 2006 (Brownlee et al., 2006) and Hayabusa, launched in 2003, returned dust grains from Itokawa (Yoshikawa et al., 2010).

Few bodies have been investigated *in situ*. The thick atmospheres of Venus, Jupiter, and Titan were studied through atmospheric probes, whereas the thin but spatially hugely extended atmospheres of comets were investigated *in situ* by selected instruments capable of performing *in situ* measurements on flyby missions. For Venus and Titan, also the surface could be reached. The first landing on Venus took place in 1975 (Veneras 9 and 10). Almost 30 years later, the Huygens probe successfully reached the surface of Titan to return the first images of this unusual moon. After the first landing on Mars in 1976 (Vikings 1 and 2), the up-to-then tremendous interest in this planet (21 missions in 15 years) stagnated for about 10 years. Subsequently, Mars investigations suffered severely from a number of mission failures (six of nine missions failed between 1988 and 1999). However, Mars has become a prime target for comparative planetology and the search for the roots of life, and Mars exploration is nowadays regarded as one of the key elements in solar system studies (e.g., Hauber and Neukum, 2006). Mars is after the Moon within our solar system the most visited place, where the extended *in situ* investigation with landers and rovers is initiated.

The exploration of any solar system body usually includes investigations with global coverage, as well as dedicated *in situ* analysis and sample return from specific regions of particular importance. This general scheme is normally being followed for the first 'round' of space missions to a solar system body, although owing to technological constraints, their exploration has practically always started with flyby missions before global characterization from an orbiting spacecraft could be achieved. At present, the only object for which all the consecutive steps mentioned earlier have actually been performed is the Moon. However, even though several hundred kilograms of Moon matter have been returned to Earth for in-depth analysis, we recently have begun a return to the Moon for global mapping, even without landing, some 35 years after the lunar sample returns. Why? Actually, we are going back for various reasons. First, it has to be verified whether the results of the detailed physical and chemical analysis of the lunar samples in the laboratory indeed apply globally to the Moon and do not just reflect some odd behavior of the particular regions, for which samples were returned. Second, a lot of progress was made over the past 35 years in modeling the formation and evolution of

solar system bodies in terms of their accretion and subsequent alteration through endogenic and exogenic processes. These models have not yet been verified through detailed measurements, largely because appropriate instruments were not available in the 1970s. Finally, a lot of new knowledge has been gathered in recent years concerning, for instance, the presence of volatiles in polar regions. All these issues can only be addressed by returning to the Moon with innovative instrumentation. To fully understand how the existence of the Moon constrains our models for the formation and evolution of the terrestrial planets, we need knowledge on its global composition, interior structure, and thermal properties, such as the surface heat flow.

By now, Mercury, Venus, Mars, Jupiter, and Saturn have been explored from orbiting satellites; Uranus and Neptune were investigated in part through flyby missions, while Pluto is still awaiting its first encounter in July 2015. Messenger is orbiting Mercury; BepiColombo is under development (Benkhoff, 2013). Rosetta is approaching 67P/Churyumov–Gerasimenko to land on the comet nucleus. While we have had a number of asteroid and comet flybys, so far, a close study from orbit has only been taken on Eros and on Itokawa by the NEAR and Hayabusa missions, respectively. Hayabusa returned around 1500 grains from asteroid Itokawa, and the Dawn mission visited Vesta and is on its way to Ceres. Venus, Mars, Jupiter, and Titan have been investigated with *in situ* probes.

Of course, we all understand that due to financial constraints, only a very limited number of space missions can be conducted at all in the next few decades. It is therefore of utmost importance to plan these mission well ahead in terms of which targets need to be visited, in which time frame, and what measurements have to be obtained of any specific target in the near and long-term future. To be able to take these important decisions, one must review the results of past missions, with a view to establishing which questions have been answered and which have remained unanswered or have not yet been addressed and why. We then have to find the appropriate strategy with a view to covering as many of the important open issues as possible within the financial and technological constraints. This approach highlights in many cases that the scientific return from past missions simply reflects the technology limitations of their time. This is particularly valid with respect to sensitivity, as well as spatial and spectral coverage of remote-sensing instrument and the availability of *in situ* measurements, suggesting that progress in scientific research goes hand in hand with the progress made in the technology of scientific instrumentation, at least for space missions.

10.23.3 Exploration Strategies and Associated Techniques

The exploration of the different types of solar system objects involves different mission strategies, as well as matched instrumentations. This, however, does not mean that a dedicated mission is required every time to one target only. Missions have been designed to combine whenever possible investigations of several objects and to cover various aspects of the chosen targets. The Galileo mission flew past two asteroids before it had its encounter with Jupiter, and Rosetta did the same before encountering its target comet. The most famous example of taking advantage of rare, but extremely fortunate, geometric constellations in the solar system is Voyager 2 using the once-every-189-years alignment of the outer planets, which allowed this spacecraft to visit all outer giant planets. However, although science can greatly benefit from spacecraft visiting various targets by the side, the mission and payload were always optimized for the investigation of the main target, and in most cases, the additional scientific return was mainly limited to images of any additionally visited body.

Future exploration strategies should therefore take advantage not only of fortunate geometric circumstances but also of advances in technology of instrumentation, which includes sensor development, as well as low-resource approaches for support and control electronics. In the following sections, some considerations are presented for shorter-term and longer-term strategies for the exploration of the inner and outer solar system.

10.23.3.1 Exploration Strategies for Near-Future Perspectives

At this point in time, planetary space exploration includes missions to Saturn and Titan (Cassini/Huygens), Mars (Mars Express, Mars Reconnaissance Orbiter, Mars Exploration Rovers, Phoenix, and Curiosity), Venus (Venus Express), Mercury (Messenger and BepiColombo), Jupiter (Juno, Juice), the Moon (Smart-1, Chandrayaan, Chang'E, and Grail), and various comets (Rosetta and Stardust) and asteroids (Hayabusa and Rosetta). Some of these missions actually represent the realization of the 'next step' in a strategic exploration approach as defined earlier. For instance, Rosetta represents the 'next step' after cometary flyby missions; Cassini represents the return to the Saturn system after Voyager, with dedicated payload and with the Huygens probe for detailed *in situ* investigations of the Titan atmosphere. When reviewing the current and past missions in relation to the status of a repeated four-step exploration approach as outlined in Section 10.23.2 (flyby, orbiter, lander, and sample return), it becomes clear that for most targets (Mars, Venus, Mercury, asteroids, and comets) in the near future, missions will need to concentrate on further *in situ* investigations from landers (atmospheric entry probes in case of Venus) and the preparation of sample return missions. After Voyager and Galileo, it is obvious that we need to return and explore the Jovian system to obtain a clear understanding of Jupiter and its magnetosphere (Juno), as well as the Galilean moons (Juice). At some point, a decision is needed on which targets to concentrate on and what type of missions to design. Three scenarios are in principle competing candidates for near-term future missions:

- Exploration of the Jovian system
- Landing on terrestrial planets
- Sample and return from low-gravity environment

We will briefly introduce possible strategies for these three types of planetary missions.

10.23.3.1.1 Exploration of the Jovian system

A staggered approach is envisaged with the final goal of having explored the major aspects of the Jovian system in a detail that

allows us to understand the formation and evolution of the system as a whole, as well as of its individual components. The Juno mission (Bolton and The Juno Team, 2012), launched in 2011, is the next step in further exploring the Jovian magnetosphere and the gravitational field, followed by ESA's Juice mission, which is in preparation for launch in 2022 (Grasset et al., 2012). Owing to the hostile environment, particularly with regard to radiation, it is difficult to investigate the inner Galilean moons over a long time. The reasonable lifetime of a Europa Orbiter is limited to approximately 60–100 days once in Europa orbit owing to the extremely harsh radiation environment. Juice took the solution to have flybys at Europa, followed by a complex tour with many flybys at Callisto, and finally an insertion into an orbit around Ganymede, the largest moon in the solar system. Remarkably, both spacecraft (Juno and Juice) are solar-powered, avoiding nuclear power generators.

Following Galileo, a very interesting scientific contribution would be provided by a new Jovian entry probe, capable of penetrating the denser layers of the atmosphere, if possible up to a pressure of 100 bar. The Galileo probe entered a dry region of the Jovian atmosphere, preventing the assessment of critical key measurements such as representative O/H ratios. To avoid duplication, a different entry strategy is required, such as entering at higher latitude, and preferably with multiple probes. The orbits required for a high-latitude entry can, however, result in higher approach velocities, hence increasing the entry velocity of the probe. This will increase the heat load on the thermal protection structure, which is already stretched to the limit of current capabilities for equatorial entry of the Jovian atmosphere.

On the longer perspective, landing on the Galilean moons (Europa and Ganymede particularly) would provide a great opportunity for surface *in situ* investigation and potentially subsurface access (by means of drilling or melting probes), though technically challenging due to the high delta-v involved for transfer to the target and the landing on the surface. An alternative solution could be the deployment of hard penetrating probes, released by a flyby or orbiting spacecraft. But accounting for the high impact velocities and related high impact shocks, the possible deployed payload would be rather limited. Further advances in miniaturization of high g-load-resistant technologies could open this perspective.

10.23.3.1.2 Preparation of landing on terrestrial planets

Depending on the presence and the thickness of their atmospheres, the landing on the terrestrial planets requires different strategies. For atmosphereless bodies such as Mercury and the Moon, the decent and soft landing must rely solely on propulsion systems and air bags. For landing on Mars and Venus, on the other hand, the descent goes through the atmosphere, which requires an atmospheric entry shield. Closer to the surface, the deceleration process can be supported by parachutes and the touchdown by air-bag systems.

Mars has become the target of prime interest for landing and *in situ* exploration. In preparation for a landing mission to Mars, two aspects need to be considered in parallel. (1) A number of improvements of the existing entry decent and landing systems will secure the safe landing on Mars, including autonomous hazard avoidance and pinpoint landing capabilities. Improved navigation accuracies before release of the entry probes, active controlled reentry with compensation of the highly variable Mars atmosphere effects during the entry, and optimization of parachutes and landing systems will further improve the safety of the landing and allow for an increase of the scientific payload deployed on the surface and access to locations yet not reachable. (2) As the landing on Mars shall be the first step toward a sample return mission, it is of utmost importance that a large area around the landing site is studied. A rover equipped with a deep drill and with science packages that include biochemical and geophysical instrumentation is required. ExoMars (under preparation for 2016 onward) and Curiosity (successfully landed in 2012) are piloting this step.

The Moon would remain as an excellent test bed for the velocity-controlled pinpoint landing system and also for the performance verification of the mobile elements and science packages. Operational sequences could be tested and improved. It has the major advantage of being very close to Earth, which reduces the travel time (and therefore complexity and cost) and simplifies communication and commanding loops. A landing system leading to a safe pinpoint landing including hazard avoidance on the Moon can provide a solid heritage for pinpoint landing on Mars, of course apart from the additional entry system and the parachutes, which decrease the required thrust proposed by a propulsion system. The mobile surface elements and science packages can be fully tested on the Moon before being sent to Mars, or even Mercury.

10.23.3.1.3 Sample and return from a low-gravity environment

Small solar system bodies, such as comets, asteroids, and TNOs, are believed to be relics of the formation of our planetary system. The near-Earth objects (NEOs) are very attractive targets for sample returns missions, because the technological challenges are less demanding than for the high-gravity environment of a planet or the volatile composition of a comet. Many NEOs can be reached on much shorter timescales than Trojan asteroids, centaurs, or TNOs. An additional benefit offers the COSPAR planetary protection classification (Category V) with a nonrestricted return from NEOs (cf. Romstedt et al., 2008). No complicated sample sealing is required to protect Earth from a potential contamination, as required for Mars or Europa sample return missions.

Dynamic calculations have shown that their orbital lifetime is small on cosmogonic timescales and that they are continuously being replenished from other sources, either from the main asteroids belt or from comets. A sample return mission needs to return a scientifically significant sample of the surface material of one or multiple NEOs. The amount of material will influence the research that can be performed. It would be advantageous to have enough material to apply all desired measurement techniques and tests. Most laboratory instruments require a very small amount of material ($\ll 1$ g) and only some investigations need a few grams. The sampling size required for one set of measurements is therefore of the order of several tens of grams. However, some redundancy is required and some additional material should be kept for future analysis. Furthermore, a larger amount would provide a better representation of the sampled area.

Retrieving a sample from a low-gravity body is significantly different from a sample return from a planet. The latter

requires a dedicated lander, an entry and/or descent system, and a launcher to return to orbit. For a low-gravity environment, no dedicated launch vehicle is required after sample collection, and also, the descent requirements are very different. A sample return from a NEO, a main belt asteroid or one of the Martian moons, can be achieved with a relative small spacecraft that can be launched on a 'low-cost' launcher, if the relative velocity between the spacecraft and the target body is low enough to permit a rendezvous mission. Upon arrival, the spacecraft will first go in orbit to perform a number of remote-sensing measurements of the surface and subsurface, as well as an accurate mapping of its gravitational properties, which will be a key prerequisite in the detailed determination of the navigation sequence for sampling maneuvers. A large number of potential sampling sites on the surface need to be examined in order to select the optimal locations for sampling maneuvers. For this, high-resolution imaging in the visual and near infrared (IR) is the minimum requirement; however, depending on the available mass, other instruments should be added to determine the global elemental and mineralogical surface composition of the target, as well as some information on the subsurface.

There are two main options to retrieve a sample from a low-gravity solar system body: (1) collecting the sample directly from the surface or (2) creating a debris cloud of asteroid material and collecting a sample from that cloud. The Hayabusa mission to asteroid Itokawa has realized the first option whereby a touch-and-go sampling approach was used. Upon a brief contact with the surface, a projectile is fired and the debris is then funneled up a cone as the spacecraft retreats from the surface. The second option has not yet been space-proven; however, the results of the Deep Impact mission to comet Tempel 1 can be used to roughly estimate the constraints for collecting a sample of some hundred grams with this method from a nonvolatile asteroid.

The samples have to be stored and sealed into a sample canister inside an Earth-entry vehicle. To reduce propellant requirements for the transfer back to Earth, the sampling mechanism and empty tanks can then be jettisoned. The spacecraft will return to Earth, where the Earth-entry capsule will perform a direct hyperbolic Earth entry.

10.23.3.2 Longer-Term Goals for Exploration

The high-priority goal is of course a sample and return from a specific set of locations that have been previously studied by *in situ* landers or their associated rovers like on Mars. While there are many differences between a sample and return mission to Mars and that to a NEO or Phobos, some technologies such as the sample encapsulation and Earth-entry vehicle may be similar. Certainly, a mission to a NEO or Phobos may be a natural test-bed precursor to this important further step on Mars.

10.23.4 Instruments

The challenge of any scientific mission is to define the scientific objectives (what is of highest priority to be observed), refine them in detailed scientific requirements, and select the instrument(s) to satisfy these requirements. The instrument accommodation drives the spacecraft design and strongly influences the overall cost of the mission. Numerous iterations are required to fit the actual selection and implementation details within the available resources (Figure 1). Trade-offs on the type of instruments, the instrument measurement principle, and combination of instruments and level of involved new technology are needed, together with the verification that the scientific objectives of the mission are still fulfilled. Usually, many compromises have to be made in the instrument selection and accommodation. Several textbooks are dealing with the space mission design approach (e.g., Wertz and Larson, 1999).

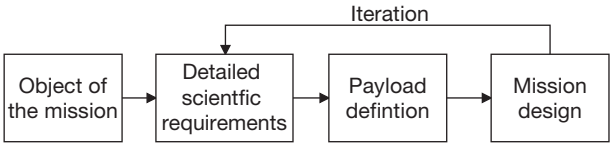


Figure 1 Definition process of a space mission.

To satisfy the objectives of the mission, a set of physical quantities have to be measured and observed by employing various observation methods. A classification of instruments according to these measurement principles should assist the reader in obtaining a structured overview of the types of instrumentation required for planetary space missions.

10.23.4.1 Measurement Principles

Planetary exploration covers a wide area from astronomy, geology, morphology, geophysics over magnetospheric and plasma science, to exobiology, just to name, but a few of the involved disciplines. A wide and ever-growing set of instruments are employed driven by new measurement requirements and new scientific disciplines on the one side and emerging new technology on the other side, enabling enhanced and new instruments. A classification of detection principles should assist in categorizing the existing range of instruments based on their measurement approach. We distinguish principles of

- direct and indirect detection
- remote and *in situ* sensing
- passive and active sensing principles

10.23.4.1.1 Direct detection

The detection principle is based on a direct measurement of the interesting physical phenomena by observation of physical quantity such as temperature (thermal sensor), pressure (pressure sensor), and particle flux by means of either remote-sensing or *in situ* observation.

10.23.4.1.2 Indirect detection

Indirect detection is performed when a direct detection is not possible or too demanding. The observation of the interesting phenomena is done by interpretation of a single or combination of several physical by-products, which are easier to access for observation than a direct measurement. Indirect detection often relies on a combination of measurements performed with a set of instruments and very often involves modeling

and simulation to derive the physical result of interest. Many processes and physical phenomena have to be measured indirectly because direct access is not possible due to demanding or inadequate space-qualified sensor systems.

Tracing of radon as product of radioactive decay of ^{238}U to conclude on volcanic activity on planetary surfaces is an example of an indirect observation of geophysical phenomena (although in many cases, volcanic activity could be detected directly by imaging, e.g., provided that visibility is given and no opaque atmosphere like on Venus or Titan blocks the direct observation). Another example is a spacecraft performing aerobraking in a planetary atmosphere, to derive the density of the atmosphere from measured decelerations.

10.23.4.1.3 Remote sensing

Any observation without directly contacting the object in question is considered remote sensing. Mainly electromagnetic waves allow for remote detection of physical qualities either directly or indirectly. The detection of electromagnetic waves covers a wide wave-band range (submillimeter (terahertz), thermal IR, far and near IR, IR, visible, ultraviolet (UV), extended UV, to x-ray and γ-rays). The reflection, scattering, and emission of photons from a planetary surface allow for the characterization, definition of chemical composition, morphology, mineralogy, etc. Due to the propagation of these photons, it is possible to detect them remotely without direct interaction of an instrument at the origin or location of the phenomena. This is exploited with remote observation either from Earth (as far as the atmospheric attenuation allows) or outside the Earth's atmosphere with space telescopes (Hubble and JWST) or in closer vicinity with remote-sensing spacecraft (flybys or in orbit). A major limiting factor for remote sensing is the distance from the observed phenomena due to the $1/r^2$ dependence from distance resulting in a limited detectable signal strength and signal-to-noise ratio (SNR). Spatial and spectral resolution is restricted by both the photon wavelength, as well as the limited instrument apertures (size, mass, and cost of an instrument), and available signal strength. Temporal resolution is limited by the minimum required exposure time. Of course, the benefit of remote sensing is that in many cases, a wide coverage on an object can be achieved. Examples of remote-sensing instruments are (telescopic) imagers, laser altimeters, and radars.

10.23.4.1.4 In situ sensing

Being at the place of a physical phenomenon (e.g., flying or floating through a planetary atmosphere with an aerobot or being on a planetary surface with lander and rovers) allows measurement of the physical qualities directly at the location. *In situ* sensing is the only possible means to investigate phenomena, which have no emission of or interaction with electromagnetic waves (or where the propagation of waves is suppressed or strongly attenuated). Due to the vicinity very often, much-lower intensity or less-disturbed signals can be measured. In most cases, *in situ* observations are of local nature (unless a high level of mobility is provided by long-distance rovers, aerobots, or orbiting spacecraft), leading to a restricted coverage. Nevertheless, it is important to provide ground truth for remote-sensing measurements, which rely often on modeling and simulations (required due to indirect measurement principles) to derive the desired information. A further benefit of *in situ* instrumentation is the possible physical interaction with the object, for example, by use of mechanical means such as grinders or drill, allowing for access of pristine and 'unweathered' material below the surface.

A major drawback for most *in situ* instrumentation is the increased complexity of the required equipment, as the instruments have to be brought to the targeted location under specific, often very challenging environmental conditions (e.g., radiation at Jupiter or extreme temperatures at Venus and Mercury). In many cases, *in situ* observations drive the mission design and increase the mission complexity, as the instruments cannot remain at a remote location but have to be brought by landers, aerobots, impactors, etc., to the location of interest.

Some examples of *in situ* instruments are dust mass spectrometers, plasma analyzers, laser mass spectrometers (LMSs), and magnetometers. Note that the return of a sample from a planetary surface (such as during the lunar Apollo program, Genesis, or for Hayabusa) also belongs in principle to the *in situ* observation category.

10.23.4.1.5 Passive detection

Passive instruments do not provide their own illumination or stimulation, but they rely on either illumination from natural sources (e.g., by the Sun) or luminescence, fluorescence, or emission of radiation. Passive instruments do not need illumination capabilities, but are directly dependent on the illumination or emission source.

Major parts of the solar emission spectra are reflected from planetary surfaces and altered by the surface and atmosphere interaction. A relative comparison of the reflected spectra with the original impinging solar spectrum reveals many surface properties. In the case of variable illumination (as in the case of solar illumination, but dependent on the wave band), a reference measurement is required to derive the relative ratio caused by the surface interaction, whenever absolute information is required. Some examples of passive instruments are microwave sounders, radiometers, x-ray spectrometers (could be also active), and magnetometers.

10.23.4.1.6 Active detection

Here, the instrument provides its own illumination using an emitter. A signal source is used to illuminate the region of interest (footprint). The sensing part is coaligned with the illuminating beam and detects the reflected signal. The benefit is of course the independence from other sources of illumination (e.g., use on the night side of a planet) due to the well-known emitted signal characteristics (time of emission, spectral features, power, etc.). The modulation of the emitted signal (e.g., with pseudocodes) can dramatically increase the SNR and assist in the detection of the emitted signal even under noisy conditions. No direct dependence on, for example, solar aspect angle (illumination aspect) or solar variability is required with this technique. The drawback is the increased complexity and resource requirements for active illumination (mainly in mass, volume, and power). It is not necessarily the case that the emitter must be in close vicinity of the receiver, but can be on another spacecraft, lander, or even located on Earth. Some examples of active instruments are radars, laser altimeters, permittivity probes, and LMSs.

Another benefit of active detection is the elegant realization of a spectral analyzer: for illumination, a source with limited wavelength emission is swept over an extended range and a detector measures the response from a target of interest.

10.23.4.2 Imager

Imaging is used to provide spatial information about the distribution of any interesting physical quantity. Imaging can be achieved either with a scanning principle like push broom and whisk broom using a single detector or line detector or by means of a two-dimensional (2-D) detector array employing multiple sensor pixels synchronously. A scanning principle either requires a mechanical moving platform or relies on the motion of the instrument (e.g., provided by an orbiting spacecraft). Scanning is used not only in cases where the detection principle is complicated and a complex and demanding instrument setup is required but also in cases where the required pixel number is exceeding the (currently) available detector array sizes or rather high-resolution spectral information is required. The imaging quality is subject to the pointing accuracy and stability of the mechanical scanning platform or the orbiting spacecraft.

In many cases, detector arrays can be used for simultaneous measurements, simplifying or even avoiding the scanning platform, but an appropriate optic is required to produce a sharp image on the detector array.

Very often, individual image frames have to be put together to provide a global image of a large spatial area. This can become rather complicated if observation distances, aspect angles, and illumination conditions are varying between individual images.

Imaging is very important to recognize features, structures, surface morphology, and context for other instruments or simply for orientation and navigation purposes. An imaging instrument uses optics such as lenses (refractor) or mirrors (reflector) to project a focused image onto a detector. The detector converts the image into electric signals, which are digitized before they are further processed with a digital data processing unit. Color imaging requires multiple exposures of the same target through different color filters. The combination of three color-filtered images in complementary colors such as red, yellow, and blue provides natural colored images. In several applications, filter wheels have been used to place the required filter in the optical path. More recent developments, however, make use of interference filters applied directly on the detector array, which does, in principle, save mass. False color images are used to highlight mineralogical composition or other specific features within the image.

Imagers consist typically of a solid-state array based on either charge-coupled device (CCD) or an active pixel sensor (APS). In earlier times, also, direct recording on film material was used.

Light falling on a detector pixel is absorbed by a photoconductive substrate such as silicon and releases a quantity of electrons proportional to the intensity of light. The stored accumulated electric charge in a pixel represents the light level integrated over the exposure time. These charges are read out for conversion into digital data. Typical CCD and APS are sensitive over a wide wave band (including UV and IR) and can be optimized for specific wave bands by selection of the detector semiconductor material. Their sensitivity can also be increased to enable the detection of single photons at specific wavelengths.

Not all imagers require 2-D detector arrays, as explained earlier. Imaging can be done also with a single-element whisk-broom sensor, multielement whisk-broom sensor, sensor line in a push-broom arrangement, or in matrix configuration. A single line detector arranged perpendicular to the spacecraft velocity vector can be used in a scanning mode (push-broom mode). A 2-D image is built up as the spacecraft moves along its velocity vector (following its orbit), requiring appropriate spacecraft attitude stability (to avoid image distortion during the scan) and precise pointing knowledge. An optical correlator system can significantly reduce stability requirements for the satellite attitude system.

The simplest approach for image sensing is through a detector array (typically 1024×1024 or larger), where a complete image is obtained in a single 'shot.' The stability requirement is now driven by the exposure time, as the platform has to remain stable during the exposure time within one pixel. The imager optics is of either a reflecting (e.g., Schmidt, Cassegrain, or triple-mirror anastigmat (TMA)) or refractive type (teleoptic or lens triplets).

10.23.4.2.1 Wide field imager

A wide field imager (WFI) has an extended field of view (FoV) with a limited ground pixel resolution intended to give an overview over a larger terrain and hence is used to provide context information for other instruments such as high-resolution cameras or spectrometers. Planetary WFIs can be rather small and low mass as the larger ground pixel size allows for a reduced aperture and reduced focal length. WFIs are usually also used in connection with stereoscopic imaging. One of the best-known WFIs is the Wide Field and Planetary Camera 2, which was installed at the Hubble Space Telescope in 1993 (Peg, 1993).

10.23.4.2.2 High-resolution camera

High-resolution imagers are used on remote-sensing spacecraft to resolve small details (typically in the meter range or even better) on the observed body surface. In combination with filters, some high-resolution compositional information can be provided as the typically employed detector arrays (CCD or APS) are also sensitive in the IR and UV region.

CCDs provide an advantage in the case of low photon counts as they have higher quantum efficiency than APS, although more recent developments in this area indicate that APS sensors can also achieve high efficiency. One significant benefit of CCDs is that time delay integration can be easily implemented due to their construction principle. The exposure time can be increased by shifting the pixel information from one pixel to the adjacent pixel by appropriate clocking of the charge transport mechanism of the CCD against the projected spacecraft velocity to allow for accumulation of photoelectric charges over time. If the shift is chosen in the right speed (determined by the ground track velocity of the spacecraft and the focal length of the optics), the spacecraft motion can be fully compensated and hence the exposure time per ground pixel increased until either saturation is achieved or the edge of

the CCD is reached. Every individual ground pixel projection is shifted through a defined number of CCD cells, before readout.

To overcome diffraction limit, the aperture of the optics has to be adapted to the required spatial resolution. The overall mass of the instrument scales with the cube of the required aperture and is significantly more demanding than the WFI. The spacecraft attitude control subsystem has to provide appropriate pointing stability and absolute pointing capabilities to avoid blurred and smeared images, and finally, a huge amount of data has to be handled by the spacecraft data handling and the telemetry subsystems. Usually, the high-resolution channel is combined with a WFI, which provides context information. Due to the amount of provided information, high-resolution imaging requires a high data rate. Image compression has to be used carefully to avoid artifacts in the final data product and loss of desired information. The high-resolution stereo camera (HRSC) on Mars Express (Neukum et al., 2004) has a ground resolution of up to 2 m per pixel; HiRise on MRO has 0.3 m per pixel at 200 km altitude (McEwen et al., 2007).

10.23.4.2.3 Stereo imager

Stereo imaging is needed to gain topographic information and to construct digital elevation models (DEMs) of planetary surfaces. Either two camera heads mounted with a spatial separation but pointing at the same target or a single camera head is used, which images a specific area at a certain time and reimages the same area under a different aspect angle (achieved by side viewing of the spacecraft) at a revisit. The first case provides stereo information with single site visits and without spacecraft off-nadir pointing. The achievable height resolution is limited by the spatial separation of the camera heads, which is usually restricted. The second solution provides almost any required separation but needs multiple revisits at different times for stereo imaging and phases of spacecraft off-nadir pointing. A reasonable compromise between both cases is the use of a second camera, which is mounted off-nadir pointing.

In all cases, complex image processing is required to retrieve the elevation of the terrain and to create 3-D images. Cameras with stereo capabilities are on board the orbiters of most planetary missions with the currently most impressive one being the HRSC on Mars Express (Neukum et al., 2004).

10.23.4.2.4 Scanning platform

Scanning platforms are used to provide flexible pointing of optical instruments independently of the spacecraft's attitude. However, most modern spacecraft designs do not include such scan platforms in order to avoid moving mechanisms, which are potential for failure. The alternative is to rotate the entire spacecraft so as to point the optical instrument in the required direction. This is particularly easily done with three-axis stabilized spacecraft, which use reaction wheels for their attitude control. The repointing can be performed without propellant use (neglecting the wheel offloading, which is required from time to time to keep the rotational rate of the reaction wheels within a specified range).

10.23.4.2.5 Panoramic camera

A panoramic camera is a high-resolution stereo imager used to image the surface and surroundings from a lander or rover. Two camera heads are spatially separated and typically mounted on top of a mast (a kind of scanning platform), allowing for an elevated view. A mechanism with two degrees of freedom permits the cameras to rotate by 360° azimuth to obtain a panoramic view and provides pointing at various elevations.

Images can be used to create 3-D maps of the area where, for example, the rover has to navigate. Mechanical filter wheels in the optical path, or fixed linear or interference filters directly mounted in front of the detector, offer multispectral imaging capabilities and color imaging. Images taken at various wavelengths are used to study mineralogical composition of interesting rocks and soils. Examples for such systems are on the Mars rovers Pathfinder, Spirit, and Opportunity (e.g., Bell et al., 2006), as well as on the Rosetta lander Philae (e.g., Willnecker et al., 2006).

A difficulty, however, arises from the large volume of data and the associated limitations given by the proximity link via an orbiting relay spacecraft. In most of the cases, a direct data transfer to Earth from the surface element is not possible.

10.23.4.2.6 Microscopic imager

The microscopic imager is a combination of a microscopic optic and a camera head. It provides images of small-scale features of rock and soil, reveals the fine-scale texture, and resolves individual grains targets, which are often further analyzed by other instruments like alpha particle x-ray spectrometer (APXS), LMS, Raman spectrometer, and Mössbauer spectrometer. Such cameras can provide multicolor images with a typical spatial resolution on the order of 50–100 μm per pixel. Illumination devices are used when night images or images in shadow regions are taken. Color information can be provided either by active illumination with light-emitting diodes at different limited wavelengths or with wideband illumination and filtering. Modern microscopic imagers can have a mass in the order of 100 g, with a power consumption of about 1 W. An example is the ROLIS imager on the Rosetta lander (Mottola et al., 2006).

10.23.4.3 Spectrometer

Spectroscopy of electromagnetic waves is a primary tool of remote sensing. It is the separation of electromagnetic waves into its component wavelengths within the full range of the electromagnetic spectrum starting from very low frequency (VLF) in the radio frequency (RF) band with wavelength of hundreds of kilometers to extreme high frequency (EHF) with several centimeters wavelength. The microwave region covers wavelengths between 0.1 and 30 cm; higher frequencies are in the millimeter and submillimeter wavelength region. The IR and visible range covers 10^{-3}–0.4×10^{-7} m, followed by UV range from 0.4×10^{-7} down to 10^{-8} m. Extremely short wavelengths are expressed in angstroms (1 Å $= 10^{-10}$ m). x-Rays are covering 100 Å down to 1 Å and below 1 Å are high-energy γ-rays.

The various spectral regions reveal specific features: radio waves for mapping gross surface properties, UV wavelengths for atmospheres and magnetospheric ions, and the near IR for understanding of mineralogical composition of solid surfaces. x-Ray and γ-rays provide data on the surface and near-surface

elemental abundances. Spectrometers measure signal intensities at individual wavelengths. Spectral information can be provided by grating, interferometric principles, filtering, heterodyne mixing, or illumination with spectrally limited and variable emitters. The spectral information can be used to reveal chemical and elemental composition, distinguish isotopes, and characterize energy distributions or to define specific atomic energy levels. There are many applications requiring spectral analysis. Of course, spectroscopy is very often combined with imaging to provide information on spatial variations and context.

The wide extent of the electromagnetic spectrum requires different detector types:

- Antennas and electromagnetic coils for the RF and microwave bands
- CCDs, APSs, microchannel plates (MCPs), and bolometer for the IR, visual, and UV bands
- Various semiconductor detectors, MCPs, and scintillators for x-ray and γ-ray

Also, the optics for the various bands is very different:

- Parabolic antennas (reflectors) and antenna groups are required for the higher-RF bands, as well as patch antennas.
- Refractive and reflective optics in the visual, IR, and UV bands.
- Microchannel optics for x-rays is a more recent development.

For high-resolution spectroscopy, the instrument stability is a key issue. Thermal stabilization of the detector and readout is often necessary. Cooling is required for low-noise applications. Calibration sources are required to eliminate instrument drifts. Sensitive detectors with very low readout noise and long observation times are required for faint objects or high-resolution spectroscopy.

10.23.4.3.1 Visual and IR mapping spectrometer

Visual and IR mapping spectrometer provides mineralogy and elemental composition in the visible and near-IR bands (around 0.7 μm to several micrometers). Improved performance can be achieved by extending the wavelength range into the thermal IR (up to around 60 μm). In the surface reflection, spectral features occur as minima due to volume scattering bands or Christiansen features and as maxima by Reststrahlen peaks and transparency features. IR mapping spectroscopy (IMS) can be used to map the surface minerals and their abundances, identify the phases and mixtures of the surface minerals, and map the mineral distributions and their correlation with images in the visible band. In addition, such an instrument can also determine the atmospheric cloud structure and its layering, the temporal and spatial variations of the constituents of the atmosphere, and the temperature versus altitude profile and define water vapor content, as well as the abundance of dust. The instruments visible and IR thermal imaging spectrometer (Coradini et al., 1999) and mercury thermal IR spectrometer (MERTIS) (Hiesinger et al., 2010) are examples for near-IR and thermal IR imaging spectrometers.

The IR light passes a baffled instrument aperture toward an optical lens or mirror system (TMA) providing the required mapping resolution. An optical filter might be applied to reject the solar flux outside the band of interest and an entrance slit in the optical path to reduce stray light. A wave-dispersing element, which could be a beam splitter cube, a prism, a variable linear filter, a grating, or a Fourier transformation (FT) concept, is used to create and project the spectral image onto the detector array. Detectors such as InGaAs, CdTe, and HgCdTe form the focal plane assembly, which requires active thermoelectric or passive cooling. New developments in the field of uncooled bolometers allow for operation up to room temperatures but still however require thermally stabilized focal plane assemblies. Typical achieved instrument spectral resolutions are of the order 20 nm or less. A tiltable mirror in the optical beam is used to alternatively expose the instrument to a calibration source.

To extract reliable compositional information from IMS spectra, a thorough understanding of the spectral features from laboratory spectroscopy on analog materials is required. Of course, local ground truth, which can be achieved through *in situ* instrumentation on landers and rovers, is very important so as to allow for the definitive interpretation of global remotely captured spectra.

10.23.4.3.2 UV spectrometer

UV spectrometers are used to determine surface composition and structure of atmospheres, the loss rates of volatile gases, and compositions and dynamics of exospheres and to search for surface layers of ice. The reflective properties of natural satellite surfaces in the UV are useful to determine the composition and physical state of the materials that comprise the surface and hence complement measurements performed with IR spectroscopy and elemental compositional information obtained from x-ray and γ-ray spectrometers. The range from 70 to 800 nm contains major resonance lines of many elements and molecules such as Ne, Ar, Na, S, Mg, Al, Fe, OH, CO_2, and H_2, even though it is difficult to cover the entire range with a single grating and detector.

For atmospheric limb observations, the level of stray light is of major concern as it decreases the SNR of the measurement. Scanning of the different atmospheric layers during limb observation is done with a movable mechanical scanning mirror. A slit in the optical path must be used to suppress the strong background of the planetary disk. A compact instrument design can be achieved with a common Rowland spectrometer configuration. Current instrument designs use third-generation holographic gratings (Singh, 1992), which produce the UV spectra on a photon-counting MCP through a planar MgF_2 window in front of the MCP. KBr or CsI photocathodes are used as detectors for the EUV and CsTe or RbTe photocathodes for the FUV and common CCD or APS imaging array for the UV and visible range.

10.23.4.4 Radiometer and Polarimeter

Radiometers provide precise intensity data by simply establishing the brightness (or the intensity of energy radiated) from targets. Polarimeters are optical instruments that measure the direction and extent of the polarization of light. They consist of a telescope fitted with a selection of polarized filters and optical detectors. Careful analysis of polarimeter data can infer

information about the composition and mechanical structure of the objects reflecting the light, such as the various chemicals and aerosols in atmospheres, rings, and satellite surfaces, as these all reflect light with differing degrees and types of polarization. Of course, a polarimeter may be integrated into another instrument, such as a camera or a photometer.

Both instrument principles require calibration sources when absolute measurements have to be provided.

10.23.4.4.1 Thermal IR radiometers

A thermal IR radiometer measures the intensity of IR energy radiated by a target, in order to determine the atmospheric and surface thermal properties and in particular the thermal inertia. When the instrument's FoV completely includes the disk of a planet, its total thermal output can be determined and as a consequence the planet's thermal energy balance derived, which in itself reveals the ratio of solar heating to the planet's internal heating, provided the latter is large enough. Thermopile arrays or bolometers are used to measure accurately parts of the Planck curve so as to establish the surface temperature. Note that the blackbody curve peak lies between several micrometers for hot bodies (e.g., Mercury) and 60–70 μm for very cold bodies (e.g., Europa). The Galileo photopolarimeter/radiometer experiment measured thermal radiation in five spectral bands between 15 and 100 mm (Hiesinger et al., 2010; Russell et al., 1992; Spinhirne et al., 2001).

10.23.4.5 Nephelometer

A spaceborne nephelometer is an instrument for measuring suspended particles (aerosols) in an atmosphere to obtain information about particle properties by scattering light in a volume containing many particles (**Figure 2**). The instrument consists in its simplest form of a light source and a light detector, which is mounted offside the direct light beam. The aerosols in the light beam produce scattering, which is measured by the detector. The light reflected into the detector is a function of the particle density, shape, color, and reflectivity. In the simplest version, only the backscattering coefficient is defined. In a more complex version, the instrument measures the elements of the scattering matrix (the matrix that relates the Stokes parameters of the scattered to the incident light in the far-field limit) (Bohren, 1983). The ultimate goal is to obtain the size distribution, particle density, and refractive indices while avoiding ambiguities caused by nonsphericity of the particles. Newer developments are based on polarimetry, as the advantages are (1) increased sensitivity to microphysical properties, (2) higher achievable accuracy, and (3) angular features insensitive to multiple scattering (Castagner and Bigio, 2006).

A light source, typically a laser operated at various wavelengths, is used with a photoelastic modulator made of a piezoelectric transducer coupled to a block of amorphous quartz to produce a periodic retardance. The incident laser light is linearly polarized and then passes through the modulator. The polarization state of the emerging light alternates between left and right circular polarization. Both the DC and the AC components of the detected scattered signal have to be measured to determine the important elements of the scattering matrix and therefore finally derive the characteristics of the suspended particles.

10.23.4.6 x-Ray Spectrometer

An x-ray spectrometer is used for remote sensing, as well as for *in situ* characterization of elemental composition through the detection of the secondary x-ray photons energy (dispersive) or photons wavelength (diffraction). Various principles are used as described in the following sections.

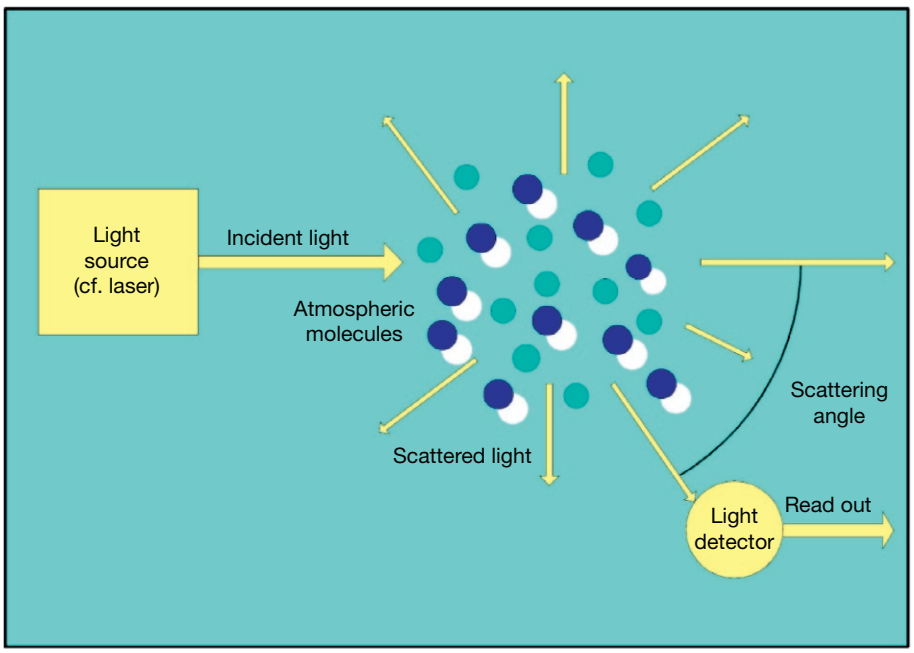


Figure 2 A nephelometer analyzes the scattering of a light source in the gas to define molecular constituents.

10.23.4.6.1 x-Ray fluorescence

x-Ray fluorescence (XRF) is based on the observation of x-ray photon emissions of atoms, which are induced into an excited state by irradiation with x-rays (e.g., originating from the Sun) causing the removal of a core electron (K, L, or M shell) from the atom. The vacant inner shells are filled by electrons from outer shells of the same atom. The difference in energy between the two electron orbitals appears as an emitted x-ray photon, which can be measured with an x-ray spectrometer either by wavelength dispersion in a diffractometer (x-ray diffraction (XRD)) or by energy dispersion with a semiconductor sensor (XRF). In both cases, this leads to information on the mineralogical composition through the determination of major-, minor-, and trace-element abundances (Meyer et al., 1996).

The XRD technique has an excellent resolution where XRF is so far more often used in space because of the smaller mass and reduced complexity. XRD instruments for space application are under development for ExoMars and Chemin (Blake et al., 2009) is successfully operating on the MSL rover Curiosity.

In an imaging XRF, a micropore optic based on a classical Wolter-I geometry (coaxial, confocal conic sections are used to form grazing incidence imaging systems as proposed by Wolter in 1952) produces a focused x-ray spot on a detector pixel array (Bavdaz et al., 2002). Semiconductor sensors based on GaAs allow for even room-temperature operation.

Remote-sensing applications of the XRF technique rely on the irradiation of the target surface with x-rays for the surrounding environment (e.g., the Sun). Solar x-ray monitors are used to observe the solar x-ray emission and to remove its spectral dependence from the XRF measurements. This is essential due to the high variability of the illuminating solar x-ray emission and the time variation and spectral hardening of the solar spectrum above 3 keV.

In situ instruments also use high-energy incident radiation sources such as an α-particle, proton or electron sources, and primary x-rays from the radioactive decay of radioisotopes such as ^{55}Fe, ^{109}Cd, and ^{238}U (Clark et al., 1977; Surkov et al., 1986). This technique can therefore be used at night or in shadow regions and thus does not depend on an external source of radiation such as the Sun.

10.23.4.6.2 Particle-induced x-ray emission

Particle-induced (or proton-induced) x-ray emission uses bombardment with ions of sufficient energy (usually MeV protons) causing inner-shell ionization of atoms in the sample. As described for XRF, outer-shell electrons drop down to replace inner-shell vacancies and x-rays of a characteristic energy are emitted. An energy-dispersive detector is then used to record and measure these x-rays and the intensities are then converted to elemental concentrations (e.g., Isabelle, 1994).

10.23.4.6.3 Alpha particle x-ray spectrometer

The APXS is used to determine *in situ* the chemical composition and absolute abundances of elements at the outermost few micrometers of soil samples (Gellert et al., 2006; Rieder et al., 1997). Also, α-particles emitted from radioactive gases (such as radon and polonium) leaking out from a planetary interior can be detected.

The APXS has evolved from the simpler α-particle instrument (APS), which uses an α-source such as Cm^{242} for target irradiation. The energy distribution of the backscattered α-particles is measured to identify light elements (except hydrogen) including rock-forming elements like Na, Mg, Al, and Si (Meyer et al., 1996). When a backscattered α-particle hits the silicon wafer detector, a small track of charge is created, which is sensed with a charge-sensitive amplifier (CSA). The charge produced is directly proportional to the energy of the backscattered alpha particle. By measuring integrated count rates over energy distribution, the elemental composition of the sample can be defined.

An improved version of the instrument, the APXS, also analyzes the x-rays produced in the sample. A ^{244}Cm source with a typical activity of around 50 m Ci emits energetic α-particles and x-rays, which are used to irradiate the surface of the target sample area at a very close distance. Hard x-rays induce fluorescence of elements with high atomic number. Medium atomic number elements, such as Ca, are excited by α-particles and x-rays. All emitted x-rays are measured with a small solid-state silicon detector over an extended period of time, and a spectrum is determined. The energy peaks in the spectrum characterize the elemental composition. The intensity in the peaks reveals the concentration of each element. The APXS is sensitive to major elements, such as Na, Mg, Al, Si, K, Ca, and Fe, and minor elements, such as P, S, Cl, Ti, Cr, Mn, and Ni. Due to the short ranges of α-particles and x-rays, the sampling depth is less than 10 μm. The APXS typically requires long integration times of order 1–2 h per sample to achieve sufficient counting statistics. Longer measurements and cold environment (night measurements) improve the counting statistics of peaks of lower intensity from elements having a lower abundance.

The APXS is preferably mounted on a robotic arm or a rover to provide mobility and access different locations for chemical compositional analysis.

10.23.4.7 Gamma-Ray Spectrometer

Gamma-ray spectroscopy (GRS) is used for determining the surface elemental composition of planetary bodies either from orbit or directly on planetary surfaces. *In situ* measurements on surfaces are preferable since they are less prone to atmospheric effects or other disturbing contributions. Direct surface measurements can provide ground truth for orbital measurements but are in most cases limited to local measurements only.

The detected gamma rays arise from two sources: (1) gamma rays emitted spontaneously by naturally occurring radioactive elements like K, Th, and U and (2) cosmic ray-induced gamma rays emitted by elements like H, C, O, Si, and Fe. Each chemical element produces a unique set of gamma-ray lines, and gamma spectroscopy allows for distinctive identification of these fingerprints and determination of the relative abundance of a specific chemical element. The ambient cosmic-ray flux produces neutron-induced reactions on elements in the planetary surface, which in turn produce the characteristic γ-rays that are used to determine the elemental concentration. A scintillator (e.g., cesium iodide (CsI), LuAP, or $LaBr_3$) is used to convert the γ-ray into light emission, which is detected by sensitive large photodiodes. The background provided by cosmic rays and by

the spacecraft is a source of noise. Mars Odyssey GRS, for example, was mounted on a 6 m boom to reduce the S/C background for the measurements (Boynton et al., 2004). Anti-Compton shielding is used to improve the GRS spectra quality to rejection of γ-rays from cosmic-ray interactions with the spacecraft material or other unwanted sources (Meyer et al., 1996). Important instrument factors are the precision of the scintillator material, sensitivity to radiation damage, quantum efficiency of the readout photo diodes, and the capability to suppress background counts. Since the γ-ray spectrometer sensitivity increases with the square root of the integration time, the expected integration times range from several orbits for the natural radio nuclides to several months for the cosmic ray-induced events. Of course, the sensitivity is also a strong function of the efficiency and energy resolution of the instrument.

Surface compositions finally are deduced from a comparison between the γ-ray spectra emitted by an area of the planetary surface with simulations based on calculated fluxes of assumed surface composition. By applying an iterative process, it is possible to derive a chemical composition that best fits the measured in-orbit fluxes.

Examples of γ-ray spectrometers based on either Ge detectors or scintillators are the γ-ray and neutron spectrometer on Messenger (Burks et al., 2004) or the Mercury gamma-ray and neutron spectrometer (Kozyrev et al., 2006) on the BepiColombo mission to Mercury.

10.23.4.8 Neutron Spectrometer

A neutron spectrometer (e.g., the HEND instrument on Mars Odyssey; Mitrofanov et al., 2002) detects free neutrons, which are produced when galactic cosmic rays interact with planetary matter. The measured neutron flux and its energy distribution depend on the chemical composition of the surface and the moderation, which takes place in the interaction of matter with the neutrons. Hydrogen is known as good moderator of neutrons, and therefore, a main application of a neutron spectrometer is the detection of hydrogen in the upper layer of the planetary surface, buried water ice, or even liquid water.

The neutron detector is, for example, a neutron-absorbing boron-loaded plastic scintillator (e.g., BC454 and GS20 glass). When neutrons pass the scintillator, material collisions occur, causing an energy loss of the neutron and creations of ion-electron pairs. When recombination occurs, photons are emitted, which can be detected with a photomultiplier tube (PMT) or photodiodes with a high sensitivity. Coincidence and anticoincidence logic is used after the readout to separate different energy bands and flux directions. The energy of these neutrons can be measured by determining the amplitude of the scintillator's light flash. The instrument data outputs are histograms of the measured amplitudes integrated over time. Thick sheets of cadmium are used to shield parts of the detector to achieve some crude spatial information or to make the detector only responsive to neutrons having energies above a certain range. It is essential to distinguish neutrons coming directly from the planetary surface and the background neutron flux. This can be achieved by using separate detectors facing different directions in combination with appropriate types of shielding.

A detector facing in the direction of the spacecraft velocity vector encounters thermal neutrons, which travel slower than the spacecraft. An additional backward detector detects thermal neutrons coming from the planetary surface. Using the combination of geometry and velocity, a separation of background and from spacecraft originated neutrons by the planetary surface is possible. The difference in counting rates between the forward and backward detectors yields a measure of the flux of thermal neutrons.

Differentiation between the different energy bands (thermal, epithermal, and fast) of the neutrons can be established through the use of different scintillator types in combination with shielding and moderation.

10.23.4.9 Laser Altimeter

A laser altimeter (also called light detection and ranging or laser image detection and ranging (LIDAR)) transmits laser pulses (energy in the order of several tens of mJ per pulse) toward planetary surfaces with a typical pulse rate of 10–30 Hz. The instrument measures the round-trip time of the light pulse between emission and the returned surface reflection to determine the distance of the spacecraft with respect to the surface. A laser altimeter operates in a very similar way to radar technology, which uses RF pulses instead of light pulses. The laser pulses are produced, for example, with a neodymium-doped yttrium aluminum garnet (Nd:YAG), which is optically pumped with laser pumping diodes. Active or passive Q-switching is used to produce the controlled laser pulses. The laser pulse emission is in the IR for a Nd:YAG (mainly 1064 nm), but with different YAG doping emission in UV, visible, or near IR that can be produced. A laser has a very narrow beam, emitted via a transmitter optic, which introduces a designed beam divergence, which allows the mapping of surface features with much higher spatial resolution compared to radar.

An optical telescope is used to collect the received reflected laser pulse photons onto an avalanche photodiode (APD). An APD is a high-biased Si-photodiode, which is very sensitive to photons and creates a charge avalanche, which is further amplified with a CSA. A very precise time measurement between pulse emission and return signal detection is used for the precise altitude estimation. For SNR increase and reduction of false trigger (e.g., by environmental photons), a very narrow filter with several-nanometer pass window at the laser emission wavelength is placed in the receiver optical path. The MLA flying on Messenger and BELA on BepiColombo (Thomas et al., 2007) are examples of laser altimeter instruments.

Repeated measurements over consecutive orbits are used to understand the morphology of planetary surfaces and construct precise topographic maps and DEMs. Repeated measurements over the same area at various seasonal periods reveal stress and strain and flexural properties of the lithosphere, determine tectonic movements and tidal deformations, and can be combined with gravity data to study the density distribution in the planetary crust. A LIDAR instrument is highly sensitive to aerosols and cloud particles in the transmission path, which opens many applications in atmospheric research and meteorology.

The receiver of a laser altimeter can be further used as a passive radiometer measuring the radiance of the surface at the instrument-sensitive wavelength.

Single photon-counting laser altimeters emit laser pulses at much higher repetition rates in the order of 20 kHz, but at much-lower energy (typically 20–50 μJ). Due to the low photon counts of the return signal, a statistical analysis of many pulses needs to be performed to retrieve the range measurement. Planetary single photon-counting laser altimeters are currently under development and promise significant reduction in mass and power.

10.23.4.10 Radar

Several solar system objects are covered by clouds or haze, making optical imaging difficult or impossible. But their atmospheres are transparent to RF waves and hence use can be made of radio detection and ranging (radar) instruments. Radars are active instruments, which illuminate a target with microwaves allowing detailed observations at any time, regardless of weather conditions, opaque atmospheres (in UV, IR, and visual bands), or presence of sunlight illumination.

In principle, the full range of the radio spectrum (Table 1) from kilometer to submillimeter wavelength can be used, according to the application and the particular requirements. The selected frequency depends on considerations, such as required footprint size, antenna size, thermal background noise, wave interaction with the target surface and wave interaction along the transmission path, and the interaction with a potential ionosphere.

10.23.4.10.1 Radar altimeter

Radar altimeters send either radio pulses (pulsed radar) or continuous wave signals (CW radar) straight down to a planet's surface (the nadir) to measure variations in the height of the terrain. In the case of the pulsed radar, the signals are timed from the instant they leave the instrument until they are reflected back. From the round-trip time, the distance can be obtained. Actual terrain height is then deduced based upon precise knowledge of the spacecraft's orbit – derived by laser or radio ranging.

CW radar altimeter often uses frequency modulation (FM) with a relative linear ramp function (for practical reasons, a triangular frequency modulation is used). During the round-trip time of the emitted signal, the local oscillator increases its frequency following the applied ramp modulation, causing a frequency difference between the transmitted and received signal. The longer the round-trip time, the larger the frequency difference becomes. The measurement of the frequency difference determines the range to the target (see Skolnik, 1980, pp. 70–92).

Laser altimeters generally have a smaller footprint, and thus higher spatial resolution, than radar altimeters. They typically also require less power. Nevertheless, there are many applications where the use of radar altimetry is beneficial, particularly for Earth observation (Gommenginger et al., 2006).

10.23.4.10.2 Ground-penetrating radar

Ground-penetrating radars (GPRs) use relatively low frequencies in the megahertz range, because signal penetration depth into matter primarily decreases with increasing frequency. Low frequencies require long antenna with several tens of meters length. For instance, the MARSIS instrument on Mars Express has 40 m long antennae (Biccari et al., 2002). With frequencies in the megahertz range, ground penetration in dry soil down to several kilometers can be achieved, with the drawback of poor spatial resolution due to the long wavelength. A large part of the signal intensity is reflected at the surface, but a significant fraction of the signal travels through the planetary crust and is reflected at subsurface interfaces between layers of different materials, including water or ice. Water is a strong attenuator due to its high permittivity, and hence, subsurface layers of liquid water can be relatively easily detected by GPRs. Nevertheless, the surface reflection and surface clutter make the GPR data analysis difficult. Ground penetrating can be used from

Table 1 Electromagnetic wave-band definition and radio spectrum nomenclature

Radio spectrum nomenclature

ELF (Hz)	*SLF (Hz)*	*ULF*	*VLF (kHz)*	*LF (kHz)*	*MF*	*HF (MHz)*	*VHF (MHz)*	*UHF*	*SHF (GHz)*	*EHF (GHz)*
3	30	300 Hz	3	30	300 kHz	3	30	300 MHz	3	30
30	300	3 kHz	30	300	3 MHz	30	300	3 GHz	30	300

Band name	*Frequency range*	*Wavelength range*	*Instrument application*
HF	3–30 MHz	100–10 m	Ground-penetrating radar
P	<300 MHz	1 m	
VHF	30–300 MHz	1–10 m	
UHF	300–1000 MHz	1–0.3 m	
L	1–2 GHz	30–15 cm	
S	2–4 GHz	15–7.5 cm	
C	4–8 GHz	7.5–3.75 cm	
X	8–12 GHz	3.75–2.5 cm	
K_u	12–18 GHz	2.5–1.67 cm	
K	18–27 GHz	1.67–1.11 cm	Water vapor measurements (22 GHz)
K_a	27–40 GHz	1.11–0.75 cm	
Mm	40–300 GHz	7.5–1 mm	Millimeter wave cloud radar
V	40–75 GHz	7.5–4. mm	
W	75–110 GHz	4–2.7 mm	Millimeter and submillimeter sounder

orbit (like MARSIS or SHARAD) or locally deployed on the surface as foreseen on the ExoMars rover (Ciarletti and The WISDOM Team, 2012).

10.23.4.10.3 Synthetic aperture radar

Synthetic aperture radar (SAR) synthesizes the angular resolving power of an antenna many times the size of the antenna aperture actually used. The SAR illuminates its target along the side of its velocity vectors and travels a distance in orbit while the reflected, phase shift-coded pulses are returning and collected. The SAR emits a permanent series of pulses as it moves along the orbit and the amplitude and phase of the individual returning signals are recorded.

A single pulse from the antenna is still rather broad (several degrees) because the diffraction limit would require a larger antenna to produce a narrower beam. The results from many pulses are combined using extensive onboard computer processing to mimic a synthetic aperture much larger than the length of the actual antenna used. The spacecraft's position and velocity must be known with great precision, and its attitude must be controlled tightly for the complex data processing, based on FT techniques (see Skolnik, 1980, pp. 517–529). The required complex data processing on board the spacecraft and the high produced data volume make the use of SAR for planetary mission difficult, and hence, SAR is mainly used for Earth observation missions.

10.23.4.10.3.1 Interferometric SAR

An interferometric SAR exploits the phase information of the returned pulses. Any phase difference of the return signals obtained simultaneously by two spatially separated antennas contains information about the angle from which the radar echo returned. Combining this with the distance information, the position in three dimensions of the image pixel can be derived.

If accommodation of two receiver antennas is not possible, measurements separated in time, perhaps from two different orbits, can be used to derive the information based on the phase shift. Even a terrain shift between observations (e.g., due to tides or earthquakes) is represented in the return phase difference and hence can be detected, provided the overall system resolution is sufficient and proper orbit and attitude data are available.

10.23.4.10.3.2 Polarimetric SAR

Anisotropic materials reflect different polarizations with different intensities. Some materials also convert one polarization into another. By emitting a mixture of polarizations and using receiving antennas with a specific polarization, several different images can be collected from the same series of pulses. Such images can be used as kind of color information in a synthesized image. Interpretation of the resulting colors requires significant testing of known materials.

10.23.4.11 Radio Science

Radio-science experiments are performed to determine masses and internal structures of bodies from spacecraft tracking data and to determine satellite radii and atmospheric structure from radio propagation information.

10.23.4.11.1 Celestial mechanics and the gravitational field

The spacecraft communication radio subsystem can be used to sense small changes in the trajectory of the spacecraft caused by gravitational field variations. For range-rate (or velocity) measurements, the spacecraft's radio transmitter sends a signal at a well-known stable frequency, generated by an ultrastable oscillator on board the spacecraft. Any velocity change of the spacecraft causes a frequency change of the radio signal received on Earth due to the Doppler effect. A spacecraft is pulled by the gravity when it passes close to a planet, moon, or any other massive body, causing its speed to change. The introduced change depends on the distance, on the total mass of the body, and on the internal mass distribution. Thus, by measuring the change in frequency (Doppler) of the Earth-received radio signal, the mass and internal structure of planetary bodies can be assessed. Only velocities in the direction of the wave propagation (along the ground station to satellite line) cause a frequency change; hence, observations are limited to spacecraft velocity changes along that direction. The closer a spacecraft approaches the body, the stronger it encounters the gravitational effects. But atmospheric drag or spacecraft internal perturbations limit the achievable resolution. A three-axis high-sensitivity accelerometer on board the spacecraft allows for separation of gravitational from other forces. Properly optimized detection of local subsurface mass concentrations (mascons) is possible.

By precise measurement of the transmission time from the ground station to the spacecraft and back (the spacecraft just 'echoes' the received signal with a slight frequency offset), an absolute distance from ground station to spacecraft (called range) can be derived. Plasma effects and any other propagation distortions in atmospheres (like rain, water content, or change in the total electron concentration (TEC)) along the line of transmission influence the measurements. A simultaneous dual-frequency link (e.g., X- and Ka-band) can be used to compensate for such effects and increase the resolution (Iess and Boscagli, 2001).

10.23.4.11.2 Radio propagation at occultation

The spacecraft radio signal can be used to investigate planetary atmospheres and ionospheres. This is performed with radio occultation experiments, when the spacecraft passes behind a planet or satellite as viewed from Earth.

The radio signal propagating from the spacecraft to Earth experiences refraction and scattering in the atmosphere of the occulting body. This causes changes in frequency and amplitude of the signal received at the ground station on Earth. Analysis of these changes yields information about the atmospheres and ionospheres. Unfortunately, the signal has to pass (in most cases unless there is a satellite-to-satellite radio experiment) also the ionosphere and atmosphere of the Earth and is influenced in the same way as described previously. However, the influence of the Earth's atmosphere can be estimated by means of additional measurements (water content definition and total electron content measurements, to mention just a few) and by selection of the location of ground stations at positions on Earth where the influence is minimal.

10.23.4.12 Microwave Sounder

Microwave sounders are remote-sensing instruments complementing IR measurements with (sub) millimeter band

measurements. A microwave sounder can be configured as continuum radiometer or a high-resolution spectral analyzer to measure naturally occurring microwave thermal emission from planetary surfaces or atmospheres or to remotely sense vertical profiles of atmospheric gases, temperature, pressure, and cloud ice (Gulkis et al., 2006). Numerous frequencies have to be analyzed to reveal atmospheric vertical temperature profiles or to determine the water vapor content (clouds, precipitation, and humidity) at various altitudes. Microwave observations are particularly sensitive to the reduced forms of oxygen and nitrogen (water and ammonia). This can be used to analyze, for example, abundances in the deep, well-mixed portion of the Jovian troposphere (Huestis et al., 2002). Data are collected to determine horizontal and vertical variations in abundances, circulation of the atmospheres, and the nature of convective and cloud-forming processes. Microwave sounding is therefore a uniquely suited remote-sensing technique to address key questions about planetary atmospheres (Huestis et al., 2002).

Not only the benefit of a passive microwave sounder is given by the global coverage, compared to local measurements of *in situ* probes like that of the Galileo probe, but also the technical realization is less challenging than for entry probes.

A parabolic antenna (or an antenna group) receives and focuses the natural microwave emission onto the very low-noise amplifier (LNA). Depending on the application, either several very narrowbands, for particular molecular abundance characterization, or a wider band for thermal emission characterization is used.

10.23.4.13 Particle Detector

10.23.4.13.1 Heavy ion counter

A heavy ion counter (HIC) instrument analyzes very high-energy heavy ions hitting the detector. Heavy ions observed during solar flares have, for example, been analyzed to determine the composition of the Sun. Measurements are also of engineering interest as heavy ions can cause random errors in spacecraft electronics. Heavy ions detectors use a stack of single-crystal silicon wafers, able to measure heavy ions with energies from a few up to hundreds of megaelectronvolts per nucleon. An example is the HIC on the Galileo spacecraft (Garrard et al., 1992).

10.23.4.13.2 Energetic particle analyzer

An energetic particle analyzer uses silicon solid-state detectors (SSDs) and a time of flight (TOF) detector system to measure the energy and angular distribution, composition, and stability of trapped radiation within a planetary environment. They are essential for the study of the interaction of energetic particles with natural satellites and with the solar wind to derive thermal plasma flow velocities and plasma temperatures and to examine adiabatic and nonthermal processes in the trapped radiation. They measure the number and energies of ions and electrons that exceed about 20 keV, define the direction of travel of such particles, and can determine the ion composition. For more information, see Funsten et al. (2004).

10.23.4.13.3 Neutral particle analyzer

A neutral particle analyzer (NPA) needs to discriminate charged particles from neutrals. This is done by applying a high voltage (in the order of a few kilovolts) across collimating optics, which repels charged particles. By switching off this voltage, the instrument is also able to measure charged particle (super thermal ions and electrons). The mass analysis of the neutrals is based on a TOF principle. An ultrasonic chopper located at the entrance point of the instrument provides a time masking, which allows neutrals to enter only at discrete very short time interval. The time difference between passing the chopper and the impact on the MCP or SSD is used to constrain the mass of the neutrals, which is given by the impact energy and the measured TOF. Detectors can be designed to have a spatial resolution (pixilated SSD or MCP with sector readout), which allows not only for TOF measurement but also for trajectory definition of incoming particles. NPAs are used to measure, for example, the escaping neutral gas composition and dynamics of planetary atmospheres and exospheres and their interaction with the solar wind, solar radiation, and meteor bombardment. The Rosina instrument on board Rosetta has unprecedented capabilities in determining the composition, velocities, and temperatures of both neutrons and ions (Balsiger et al., 2007).

10.23.4.14 Mass Spectrometer

Mass spectrometry is useful to characterize atoms or molecules in a sample, as well as determining chemical and structural information about molecules. Mass spectrometers use the difference in the mass-to-charge ratio of ionized atoms or molecules to separate them from each other. All commonly used mass analyzers use electric and magnetic fields to apply a force on charged particles (ions).

The general operational steps for a mass spectrometer are to

- create gas phase ions,
- separate the ions in space or time based on their mass-to-charge ratio, and
- measure the quantity of ions of each mass-to-charge ratio.

Different principles for mass spectrometer in use are

- magnetic focusing analyzer,
- quadrupole mass analyzer (QMA),
- ion trap,
- Wien filter,
- TOF mass analyzer.

10.23.4.14.1 Magnetic focusing analyzer

Magnetic sector-field mass spectrometers consist of an ion source, a magnetic momentum analyzer, and channeltron electron multipliers or MCPs as detector. Ions formed by electron bombardment are accelerated and focused by an applied electric field in the ion optics and collimated into a beam, which then traverses a magnetic field that bends it onto a collector slit. The ions that successfully traverse the curved path of the mass spectrometer and pass the slit are a function of the applied magnetic field strength, the mass-to-charge ratio, and the ion momentum. This principle is used to select different ion masses. Some spectrometers use multiple collectors, with multiple different path radii focused simultaneously on multiple collector slits, to cover a wide mass range. In this way, multiple mass peaks may be monitored simultaneously.

Behind each slit high-counting-rate channeltron, electron multipliers are counting the arriving ions. A specific ion mass can be selected with a particular voltage applied to the ion optics. Scanning through a voltage range provides a counting rate, which is a function of the applied voltage, and allows for the determination of a mass spectrum as the resulting output.

Magnetic mass spectrometers may be combined with electrostatic analyzer to achieve double focusing for both bending angle and energy, which leads to a higher mass resolution than single-focusing magnetic analyzers.

10.23.4.14.2 Quadrupole mass analyzer

The QMA utilizes four parallel cylindrical or hyperbolically shaped rods (Figure 3). The rods must be long relative to the inner surrounding diameter, to minimize fringing fields on the active length of the rods. A quadrupole potential is established by applying a time-varying potential on alternate rods with an RF (up to several megahertz frequency) and a DC component.

The ions are injected into the rod arrangement, and only those ions having the correct mass-to-charge ratio can pass to the exit aperture without colliding into one of the rods. Hence, the quadrupole arrangement works as a mass filter. A full mass spectrum is obtained by sweeping the DC offset and the RF amplitude at a fixed frequency and detecting the transmitted masses over the sweep at the quadrupole exit with a Faraday cup or particle multiplier. The resolution of the device depends on the rod geometry, frequency, rod length, and axial and radial ion injection energies. An example of a recently flown QMA is the SAM instrument on MSL (Mahaffy and The SAM Team, 2011).

Improved QMAs make use of hyperbolic rods with narrow wings, resulting in a better mass resolution. A miniature quadrupole array has already been constructed with a mass of several grams (excluding control and readout electronics). The future challenge is to miniaturize the control and readout electronics, as well as the ionization source.

10.23.4.14.3 Ion trap

A quadrupole ion trap uses three electrodes to trap and accumulate ions in a small volume (Figure 4). A middle ring electrode separates two hemispherical electrodes that are connected to ground potential. The middle ring is connected to an RF voltage with frequencies up to several megahertz with some DC offset. Ions brought into this arrangement are trapped in the RF field and perform stable oscillations between the electrodes. Ions of a particular mass-to-charge ratio can be ejected by increasing the RF signal amplitude. A full mass spectrum is obtained by sweeping the electrode voltages over the full range to eject all ions having different mass-to-charge ratio from the trap into an electron multiplier. A full spectrum can be obtained by counting the events at the electron multiplier output as function of the electrode control voltage. Ion traps have a very excellent mass resolution with good signal-to-noise behavior as the trap effectively works as an accumulator of input ions. Full ion trap analyzers can be designed with a mass of less than 100 g.

10.23.4.14.4 Wien filter

In a Wien filter, ions are injected in a perpendicular direction to a crossed electric and magnetic field. Only ions with the matched injection energy will pass this region linearly, when the electric force (Coulomb force depends on ion charge and electric field strength) matches exactly the magnetic deflection (Lorentz force is dependent on ion velocity, ion charge, and magnetic field strength). Ions with different injection energy will follow a nonlinear trajectory as the electric force does not match the magnetic force. A spatially extended detector (MCP, CCD, or multianode detector) after the Wien filter is used to

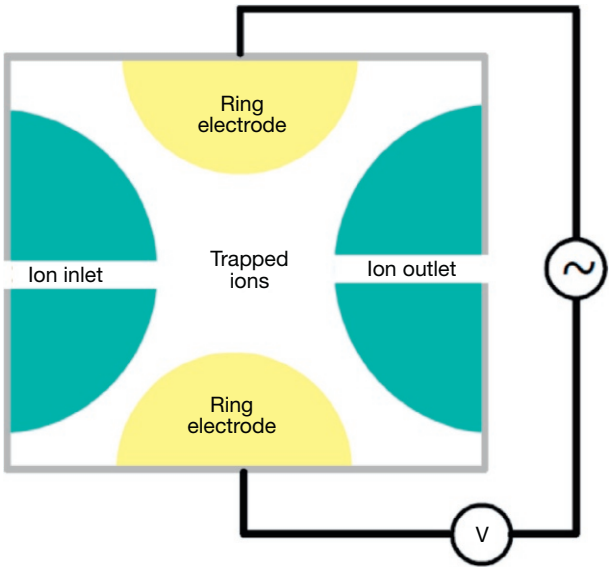


Figure 4 Principle of an ion trap.

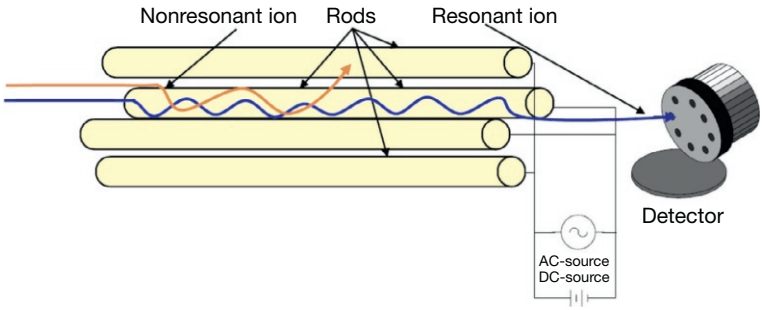


Figure 3 Principle of a quadrupole mass spectrometer.

detect different deflected ions and a mass spectrum can be derived. A Wien filter is also used to remove ions of unwanted isotopes and impurities, multiple charged ions, and neutral particles.

10.23.4.14.5 Gas chromatograph–mass spectrometer

Gas chromatography–mass spectroscopy (GCMS) is used to measure abundances and isotopic ratios of noble gases (He, Ne, Ar, Kr, and Xe) and the abundances of minor gas constituents like SO_2, COS, CO, HCl, H_2S, and H_2O (nonexhaustive) in planetary atmospheres. A pump system controlled with valves is used to generate a low gas flow rate at a reduced pressure to obtain new gas samples from the ambient atmosphere, which are combined with a carrier gas (such as hydrogen) and injected into the gas chromatograph (GC). The GC section provides a time-dependent separation of the gas components prior to a molecular identification in the mass spectrometer (MS). The output of the GC section is ionized with, for example, electron impact ionization before arriving in a quadrupole mass filter, followed by an ion detector. But other methods for the MS can be used as well.

Gas chromatograph–mass spectrometer systems are powerful analytic tools for chemical analysis of many compounds and especially of gas mixtures, but relative complex instruments. GCMS instruments have been flown, for example, on Huygens (Niemann et al., 2002) and the Rosetta lander Philae (Goesmann et al., 2006; Wright et al., 2006) or more recently on the MSL mission (Mahaffy and The SAM Team, 2011).

10.23.4.14.6 TOF mass analyzer

A TOF mass spectrometer consists of an ion source, accelerator, timing gates, a long field-free flight tube, and particle detector. The source ions are accelerated in an electric field and injected into the drift region of the TOF. Gates at the entrance of the drift tube are opened only for a short time to allow a group of ions to enter the drift tube. Ions having a different mass-to-charge ratio take different transit times through the drift region. Lighter ions have a higher velocity than heavier ions and hence reach the detector at the end of the drift region faster. To avoid mass-aliasing effects, additional gates are introduced in the drift region, before the detector, and at the end of the drift region. A TOF analyzer has a relatively large mass range and a good sensitivity provided the drift tube is not too small.

10.23.4.14.7 Laser ablation mass spectrometer

A short well-focused pulsed laser beam is used to ablate small amounts of a target surface material and to convert it into low-energy plasma. The ablation can be done even without prior sample preparation. Emerging ions are analyzed within a TOF mass spectrometer, which determines elemental and isotopic composition of the ablated material. The electrostatic fields in the reflectron of the TOF are designed to cancel energy dispersion of the ions over a broad energy range to achieve an ion TOF from the point of ablation to the MCP being independent of ion energy and proportional to the square root of ion mass. For compact instruments, the TOF ranges from a few microseconds (for H) to <100 μs (for >200 amu). The MCP current as a function of time relative to the laser pulse provides a mass spectrum of the analyzed material. The MCP output signal can be digitized with high accuracy and high time resolution, thus providing isotopic mass resolution for various elements. A laser ablation mass spectrometer was flown on the Phobos mission, LIMA-D (e.g., Pellinen et al., 1990). Modern instruments are small and of low power and low mass (250 g instruments are under development), making them ideally suited for missions to planets and small bodies in the solar system to provide quantitative analysis for major elements and isotopes present in a sample down to levels below 100 ppm. See Brinkhoff (2005) for an analysis of the application to small bodies.

10.23.4.14.8 Laser-induced breakdown spectroscopy

Laser-induced breakdown spectroscopy (LIBS) or laser-induced plasma spectroscopy is atomic-emission spectroscopy with high-energy laser pulse as the excitation source. It consists of a solid-state laser and a spectrometer. The laser emits short (typically 10 ns) focused pulses with a local power density of around $10^7\ W\ m^{-2}$, focused onto a very small point on a sample. Such high energy ablates small amounts of material and converts it into a plasma plume, which breaks down into excited ionic and atomic species. At these high temperatures, the plasma emits continuum radiation and cannot be analyzed, but the plasma rapidly expands and cools (within 10 μs) and consequently reveals the characteristic atomic-emission lines of the elements contained, which is analyzed with a spectrometer with a polychromator (Echelle type) and a CCD or APS detector for readout. The spectrometer range is typically from 1.1 μm to 170 nm covering almost all elements and within the sensitivity range of a classical CCD or APS detector.

The major advantage of the LIBS technique is its ability to provide a depth profile of a sample by repeatedly focusing the laser pulses on the same spot and analyzing the generated plasma. Spectra can be produced within seconds (compared to hours with APX or Mössbauer). LIBS also allows for some remote analysis of samples by employing telescopic optics. For proper measurements, the laser output needs to be kept stable.

A LIBS utilizes many components that are also used for a Raman spectrometer. A combined LIBS–Raman spectrometer allows for both atomic and molecular characterization of a sample. The LIBS–Raman principle is described in Giakoumaki et al. (2006). Currently combined miniaturized LIBS–Raman instruments for space applications are under development (Rull et al., 2011).

10.23.4.14.9 Raman spectrometer

Raman spectrometer (Sharma, et al., 2005) uses the inelastic (Raman) scattering of light to analyze vibrational and rotational modes of molecules. Monochromatic light, produced from a laser in the visible, near-IR, or near-UV range, is scattered by a molecule or by a crystal. One photon of the incident radiation is annihilated and, at the same time, one scattered photon is created. In the case of elastic Rayleigh scattering, the energy of the scattered photon is equal to that of the incident one. Alternatively, the energy of the scattered photon is different to that of the incident one, when inelastic Raman scattering occurs. Spontaneous Raman scattering is very weak, and separating the weak inelastic scattered light from the intense Rayleigh-scattered laser light is the main challenge.

The spectrum of the scattered light consists of a strong exciting line of the wavelength of the incident laser light due to the Rayleigh scattering together with the weaker Raman lines at different wavelengths. The individual shift in wavelength gives information about the phonon modes in the illuminated matter. The lines having longer wavelength are called Stokes lines, the others anti-Stokes lines.

In a Raman spectrometer, the sample is illuminated with a laser beam. The light from the illuminated spot is collected with a lens and sent through a monochromator. Wavelengths close to the laser emission line are filtered out (to remove the elastic Rayleigh scattering), and those in a certain spectral window away from the laser line are dispersed onto a detector. Holographic diffraction gratings and multiple dispersion stages are typically used to achieve a high degree of laser rejection. A photon-counting PMT or a CCD can be used to detect the Raman-scattered light.

10.23.4.15 Mössbauer Spectrometer

A Mössbauer spectrometer determines *in situ* the mineralogical composition of soil samples and provides quantitative analysis of Fe-bearing materials. It can determine the composition and abundance of these minerals, as well as their magnetic properties to a high level of accuracy. It makes use of the resonance absorption of γ-rays by Fe^{57} nuclei, based on the Mössbauer effect. The absorption provides information on the nuclear energy-level splitting of Fe-bearing materials. The splitting reflects the nuclear levels occupied by surrounding electrons.

Co^{57} is normally used as the γ-ray source. In general, the structure of the nuclear energy levels of the soil samples differs from that of the Co^{57} γ-ray source because of different oxidation states, chemical environment, and magnetic order. To achieve resonance with the soil samples, the energy levels of the emitted γ-rays from the source are modulated to overlap with energy levels of the sample. The Co^{57} source is mounted on an electromechanical velocity transducer and moved with respect to the target, leading to velocity-modulated energy levels of the emitted γ-rays due to the Doppler effect. Accordingly, the energy increases if the source moves toward the sample and decreases when moved away. Phase and the oxidation state of the target can be determined from the peak locations in the resulting spectrum. The concentration can be estimated from the peak intensities in this spectrum.

Analysis of material can be performed by either transmission or backscattering, although the transmission method is limited to thin targets directly placed between the source and the detector.

With the backscattering principle, only the surface of the sample contributes to the received spectrum. The average diagnostic depth for a backscattering Mössbauer spectrometer is in the order of 100–200 μm (assuming basaltic rock composition). The backscattering approach is generally preferred for planetary applications.

The Mössbauer spectrometer is sensitive to temperature variations, which should not exceed about $\pm 10\,^{\circ}C$ during acquisition of a single spectrum to minimize the temperature-dependent parameter variations and the resulting spectral smearing. Measurements are taken by placing the instrument's sensor head directly against a rock or soil sample. A Mössbauer measurement typically can take about 12 h. Mössbauer spectrometers are preferably mounted on microrovers or robotic arms to provide mobility (e.g., Klingelhöfer et al., 2003; Wade et al., 1999).

10.23.4.16 Dust Analyzer

A dust detector instrument is an *in situ* instrument used to measure the mass, electric charge, and velocity of incoming dust particles. The detectable masses of dust particles are typically in the range 10^{-16}–10^{-6} g. The speed of these small dust particles is measured over a typical range of 1 km to several tens of kilometers per second at impact rates from 1 particle per 100 days to several hundreds of particle impacts per second. The instrument aperture is the important characteristic to guarantee a proper count rate for a specific application. The length of the flight path of the TOF mass analysis section defines the achievable mass resolution and contributes, together with the aperture, to the overall instrument size. Dust compositional analysis instruments have flown on various comet missions (e.g., Kissel et al., 2003).

10.23.4.16.1 Quartz crystal microbalance

A quartz crystal microbalance (QCM) can detect tiny mass deposits on the sensor crystal by the resonance frequency change introduced by the addition of the small mass. Frequency measurements can be made with high precision; hence, it is easy to measure small mass increments with a QCM. Correlation between mass and frequency is achieved by means of the Sauerbrey equation (Sauerbrey, 1959). For compensation of the strong temperature dependence, a sealed reference quartz system is required, which is exposed to same temperature, but is protected from additional particle deposits. An example of a QCM flown in space is that forming part of the Electric Propulsion Diagnostic Package on board SMART-1.

10.23.4.17 Radiation Environment Monitor

Radiation monitoring during cruise phase and operational phase of a space mission is not only important from an engineering point of view but also scientifically of high interest. Observations of solar flares or measurements of the Earth or very strong Jovian radiation belts are some examples. A radiation monitor has to cover a wide energy range starting at several kiloelectronvolts up to gigaelectronvolts and must be sensitive to electrons, protons, heavy ions, and galactic cosmic rays. Only a combination of various detectors (SSDs, scintillator, and scintillating fiber) can satisfy these requirements. Important issues are (1) the flux measurement to derive the energy spectrum and (2) the measurement of the total (integrated) dose.

Newer technologies employ thermoluminescent dosimeters for local dosimetry and radiation-sensitive field effect transistor (metal oxide semiconductor structures sensitive to ionizing radiation) as real-time dosimeters (Bühler et al., 1996). Several new radiation monitors are under development (see Taylor et al., 2011).

10.23.4.18 Scanning Tunneling and Atomic Force Microscope

A scanning tunneling microscope (STM) is used to image surfaces of electrically conducting materials at atomic scales. An atomically sharp metal tip is brought very close to the surface, and a voltage is applied between the sensing tip and the sample, driving a small tunneling current whose magnitude depends on the distance between the tip and the surface. As the tip is moved laterally across the surface, a feedback mechanism moves the tip up and down to maintain a constant tunneling current. 2-D raster scanning with the tip across the surface produces a topographic map of the surface.

An atomic-force microscope (AFM) works on a similar principle, but is not restricted to electrically conductive materials. The AFM tip is sensitive to atomic and intermolecular forces (such as van der Waal forces) and moves up or down on its supporting cantilever. The movement of the tip is monitored with a laser beam reflected from the tip and read out with a photodiode array or by piezoresistive elements, which follows the movement of the tip and is then read out by a Wheatstone bridge with picometer resolution. Raster scanning the tip across the surface produces a topographic map of the surface with atomic resolution. Piezoceramics are used for the 3-D positioning of the tip at high precision. A major difficulty of STM and AFM for space application is the sample preparation and sample positioning in front of the tip, which requires complicated and sensitive mechanisms. The Midas instrument is an AFM flying on Rosetta (Riedler et al., 2006).

10.23.4.19 Scanning Electron Microscope

A scanning electron microscope (SEM) provides high-resolution 3-D images of a sample surface. An electron beam (few hundred electronvolts to 50 keV) thermionically emitted from a tungsten or lanthanum hexaboride cathode or alternatively by field emission (FE) is accelerated toward an anode and focused by a magnetic lens into a beam with a focal spot the size of a few nanometers. Electromagnetic coils are used to steer the electron beam in a scanning mode over a rectangular area of the sample surface.

The electrons hit the surface and are inelastically scattered by atoms in the sample. The scattering process spreads the focused beam up to a few microns onto the surface of the sample leading to emission of electrons and x-rays. The low-energy (<50 eV) secondary electrons can be detected by a scintillator–photomultiplier device, and due to the raster scan, a 2-D intensity distribution can be displayed as an image. The number of secondary electrons reaching the detector is translated into the brightness of the image. Emitted x-rays can be analyzed with dispersive x-ray spectroscopy. Also, the backscattered electrons can be detected and analyzed as an electron backscattered diffraction image to determine the crystallographic structure of the sample surface.

The spatial resolution of an SEM depends on the size of the focused electron spot defined by the electromagnetic coil system and the extension of the volume in which the electron beam interacts with the sample material. The resolution of the SEM is not sufficiently high to image down to the atomic scale. A miniaturized (11.9 kg) design of a spaceborne SEM has been developed for the (cancelled) Comet Rendezvous Asteroid Flyby mission (Albee and Bradley, 1987).

10.23.4.20 Plasma and Wave Instruments

Plasma, the most common phase of matter in the universe, consists of partially or fully ionized gas. It contains an equal number of positive and negative charge carriers and hence behaves quasineutral in the stationary state. There are quite a number of different geophysical plasmas with a wide spread in their characteristics, ranging from lower density and lower temperature to high-density and high-temperature plasmas (Baumjohann and Treumann, 1996). Basically, there are two set of instruments required for plasma studies investigating either the associated electromagnetic waves (plasma wave analyzers (PWAs)) or the charge carriers of the plasma (plasma analyzer).

10.23.4.20.1 Plasma wave analyzer

PWAs measure the electrostatic and electromagnetic components of plasma waves. The instrument receiver is sensitive to the wavelengths from several hertz to several hundred kilohertz in some case to the megahertz range. PWAs are used for a wide range of plasma investigations, for example, magnetospheres, ionospheres, solar wind, atmospheric lightning detection, or events when dust particles strike the spacecraft.

Plasma particles are bound to the natural magnetic field. Motions within the plasma will cause changes with time of the electric and magnetic fields detectable as plasma waves. An electric dipole antenna is used to measure the electric field component in the plasma, while (tri-axial) search coils are used for the magnetic field components. In the case where very low electric field frequencies have to be analyzed, long dipole antennas are required, sometimes exceeding 50 m single tip length (Gustafsson et al., 1997, 2001). Such booms are usually designed as wire booms and accommodated on spinning spacecraft providing gyroscopic stiffness to the wire booms. The deployment of such booms can be problematic and risky for the spacecraft but is by now demonstrated successfully by several space missions.

The received antenna signal is amplified with wideband electrometer preamplifiers or a set of band-limited preamplifiers for frequency-selective analysis. Low-noise design and high impedance with guard feedback systems are required for low-density and weak-field analysis to minimize the disturbances of the environmental plasma and to achieve enough sensitivity. The preamplified signals are usually further amplified for frequency analysis with either discrete filter banks or computational Fourier analysis using digital signal processing (DSP). The same DSP and analog-to-digital conversion chain can be also used for the magnetic field search coil system. Nearly simultaneous measurements of the electric and magnetic field spectrum allow electrostatic waves to be distinguished from electromagnetic waves.

Examples of PWAs are the wideband plasma wave investigation for the Earth's magnetosphere on Cluster II (Gurnett et al., 1997), the low-frequency PWA (LFA) on board the Planet-B spacecraft for Mars (Matsumoto et al., 1998), and the Huygens permittivity, waves, and altimetry

analyzer (PWA) for the atmosphere of Titan (Hamelin et al., 2000).

10.23.4.20.2 Plasma analyzer

Plasma analyzers are used to assess the elemental and isotopic composition, the associated energy levels (temperature), density, velocity, and 3-D distribution of interplanetary, planetary ionospheric and magnetospheric, and solar wind plasma. They are further used to determine the sources of the magnetospheric plasma, investigate the interaction of ambient plasma with moons and planets, and study magnetosphere dynamics, corotational forces, field-aligned currents, and magnetic reconnection and the nature of the equatorial current sheet of magnetospheres.

Plasma analyzers are in most cases combinations of instruments to satisfy the wide scope of measurements:

An electrostatic analyzer (ESA) is used to measure the deflection of the ions (and electrons) in a static electric field to define their E/q ratio (energy (E) over charge (q)).

A TOF section is used to measure TOF of the (static preaccelerated) ions and their energy with an SSD (see Solar Wind Ion Composition Spectrometer (SWICS) on Ulysses; Gloeckler et al., 1992). From the ESA and TOF measurements, simple equations can be used to estimate the energy E, charge state q, and mass m of the ions.

An electron analyzer system (EAS) is used to define the energy of electrons (typically up to a few kiloelectronvolts). It consists of two or three combined sensor heads and an ESA with electrostatic deflection (e.g., SWOOPS on Ulysses; Bame et al., 1992). See also ASPERA-3 on Mars Express, which measures ions, electrons, and high-energy neutral atoms in the outer atmosphere of Mars.

10.23.4.20.3 Langmuir probe

A Langmuir probe is used to determine the electron temperature, electron density, and plasma potential, by inserting one or more electrodes (segmented Langmuir probe) into the plasma, with a time-varying electric potential between the electrode and the surrounding plasma. The measured currents over the applied potentials (I–U curve) allow for determination of the main physical properties of the plasma. Although the measurement itself is rather simple, the interpretation of the data requires deeper understanding of the Langmuir probe theory for a spherical probe in unmagnetized plasma (Schott, 1968).

The instrument consists of a short boom (to overcome the Debye length and associated plasma distortions) with the sensitive electrode and guard system. The instrument controller applies a time-varying potential to the electrode and measures with a current-sensitive amplifier the associated current to or from the plasma (the current direction depends on the relative applied potential).

One example is the ISL Langmuir probe experiment on board DEMETER (Lebreton et al., 2006).

10.23.4.20.4 Electron drift instrument

An electron drift instrument (EDI) measures the displacement of a weak (<1 μA) artificially injected electron beam in an ambient electromagnetic field at different electron energies and beam directions. The electromagnetic field introduces a drift of the electron beam. A detector on the opposite side of the spacecraft detects the beam only at some specific geometric and energetic conditions, dependent on the ambient electromagnetic field strength, the angle between injection and magnetic field vectors, and injection energies of the electron beam. Two beams and two detectors allow for continuous direction monitoring, and by triangulation, the displacement can be derived. For small magnetic fields, the displacement is achieved from TOF difference of the two beams. Supplementary data from magnetometer measurements (fluxgate and search coil) combined with the EDI TOF and triangulation are used to constrain the ambient electric field. A key advantage of the EDI technique is that the beam probes the ambient electric field at several kilometers away from the spacecraft and hence under undisturbed conditions. As a by-product, the measured TOFs provide a precise measurement of the magnetic field magnitude (Paschmann et al., 1997).

The narrow ejected electron beam must be controlled within a hemisphere. Electrostatic multipoles are used to deflect the electron beam in the desired direction. Also, the detector has a selectable direction over a wide range. A complex optic in combination with an annular MCP is used to adjust the observational direction of the detector (Paschmann et al., 1997). A scanning algorithm employing various emission and detector directions combined with several electron beam energies and beam currents is used to find conditions where the beam returns back to the detector via the ambient field after single or multiple gyrations. The beam, produced by a tungsten filament, is pseudonoise-coded to allow for TOF analysis by correlation (Nakamura et al., 1989).

10.23.4.20.5 Active spacecraft potential control

Measurements of electric field potential are subject to distortions from interaction of the spacecraft with the plasma and by photoelectrons. Occurring spacecraft floating potentials may prohibit measurement of cold ambient plasma (at several electron volts), as low energetic ions and electrons are repelled by the potential of the spacecraft. To allow for measurements of lower energetic plasma, an active spacecraft is required. A Langmuir probe (or similar instrument) is used to measure the actual spacecraft potential, and an active ion source is employed to compensate the spacecraft charging by emission of an ion beam. A liquid-metal ion source has been used for active spacecraft potential control on Cluster II (Riedler et al., 1997): A solid needle usually made from tungsten is mounted in a reservoir with heated melted indium (at about 520 K). A potential of several kilovolts is applied between the needle and an extractor electrode, which causes the liquid metal to be pulled toward the extractor electrode. Positively charged ions are evaporated in the strong field of the needle and the hydrodynamic flow causes liquid metal from the reservoir to follow. The controlled emission of the ions is finally used to reduce the charge of the spacecraft.

10.23.4.21 Magnetometer

Magnetometers are used to map magnetospheres of planetary bodies and analyze their dynamics and fluctuations. Magnetometers are usually mounted on the tip of long booms to reduce the magnetic background introduced by the spacecraft

and its components. However, not all these effects can be eliminated by separating the instrument by a boom. An additional magnetometer (inbound magnetometer) mounted at a specific boom length fraction can help distinguish spacecraft-produced fields from ambient fields. In case of high-resolution measurements, the spacecraft has to undergo a magnetic cleanliness program, with specific design rules and material selection constraints, which introduces cost issues for the mission. Another source of potential error in measurement comes from bending and twisting of the long magnetometer boom. Increase of the boom rigidity (mass increase) or measurement of the actual bending (e.g., with a small stellar compass) can be used to mitigate the influence. In some cases, the use of an active boom bending control (smart structures) can be applied. The Ulysses magnetometer uses two sensors representing the two major measurement principles: a fluxgate and a vector helium magnetometer (Balogh et al., 1992). The magnetic fields of the planets and techniques of interpreting magnetic field measurements are discussed in **Chapter 10.07**.

10.23.4.21.1 Fluxgate magnetometer

A fluxgate magnetometer (FGM) sensor consists of three coils wound around a ferromagnetic ring core, with the primary and secondary windings aligned within the plane of the sensor core, and an additional calibration winding. The primary windings are driven by a current pulse train at a frequency f_0 such that the core is in its region of saturation. In the presence of an external magnetic field, a second harmonic (at the frequency $2f_0$) is generated in the pickup coil (secondary winding), with its amplitude proportional to the coaligned component of the external field. The phase relation between the pickup signals and the primary drive signal represents the direction of the magnetic field component.

Usually, three identical single-axis fluxgate sensors in an orthogonal arrangement are used to measure the three orthogonal components of the magnetic field. FGMs provide in general sufficient sensitivity but are subject to some long-time drift, which has to be compensated.

10.23.4.21.2 Vector helium magnetometer

Vector helium magnetometers are based on the influence an ambient magnetic field has on the efficiency with which a metastable population of helium gas in the triplet ground state can be optically pumped (Zeeman effect). A good example is the Cassini vector/scalar helium magnetometer (Dougherty et al., 2004): Metastable helium is generated in a helium-absorption cell by RF discharge. An incident radiation produces a distribution of atoms in nonthermodynamic equilibrium among the various energetic substates by optical pumping. Circularly polarized light of wavelength 1.08 μm is used for the pumping. The presence of an ambient magnetic field affects the pumping efficiency, which is exploited in the vector mode of the magnetometer: An IR detector senses the changes in absorption in the helium cell. For a constant light output, the change in pumping efficiency is measured, which depends on the magnetic field strength and angle between the magnetic field vectors with respect to the optical axis of the system.

In the scalar mode of the V/SHM magnetometer, IR radiation and an AC magnetic field are simultaneously applied to the absorption cell. The efficiency of the optical pumping is minimal, and the absorption is greatest when the frequency of the applied AC field is at the Larmor frequency (precession frequency of the magnetic moments of electrons, atomic nuclei, and atoms around the direction of an external magnetic field), which is directly proportional to the ambient magnetic field. Vector helium magnetometers are very sensitive and tend to less drifting compared to FGM (Dougherty et al., 2004).

10.23.4.21.3 Search coil

Search coil magnetometers are based on Faraday's law of induction and used to measure the AC magnetic field. Several turns of conducting wire are wound around a highly permeable ferromagnetic material core. A varying magnetic field passing through the coil induces a voltage, which is proportional to the time variation of the magnetic field. The sensitivity of search coil magnetometer directly depends on the area of the coil, the number of turns, the permeability of the core, and the rate change of magnetic flux. Specially designed low-noise and low-power amplifiers are used to read out the induced voltage from several hertz to several kilohertz (Coillot et al., 2006).

10.23.4.22 Accelerometer and Seismometer

Accelerometers are used to sense acceleration of vehicles such as rovers, aerobots, descent probes, landers, and spacecraft to

- derive atmospheric parameters (density and drag) during descent/entry phases,
- detect spin rate,
- derive nongravitational perturbations acting on the spacecraft to determine the actual gravitational field in combination with radio-science experiments,
- investigate seismic events, and
- perform inertial navigation.

During the entry and descent phase of an entry probe, the interaction with the atmospheres causes drag on the probe, leading to a strong deceleration, which can be measured with accelerometers. Out of the measured data, an estimation of the atmospheric density can be derived (Fulchignioni et al., 2002).

For the entry probe stabilization during the descent, a spin is often introduced. The spin rate can be measured with accelerometers. This is, besides the engineering aspects, of importance as auxiliary information for scientific measurements performed during the descent (Lorenz, 2004).

The direct solar radiation and the thermal emission from the planetary surfaces exert a force upon an orbiting spacecraft that, if not accounted for, destroys the accuracy of orbital parameter determination (cf. spacecraft tracking). Very sensitive three-axis accelerometers are used on board spacecraft to separate these nongravitational from gravitational components (Iafolla and Nozzoli, 2001).

Various types of accelerometer exist based on the following main principles:

- Piezoelectric effect
- Capacitive effect
- Microelectromechanical systems (MEMS)
- Electromechanical servo principle

Piezoelectric accelerometers use a proof mass, which causes a shear stress to the sensing piezoelectric crystals due to the

acceleration. This stress results in a proportional electric output signal of the piezoelectric material. With this technique, rather simple accelerometer can be constructed. A drawback is the thermal sensitivity of the crystal material, which requires an appropriate thermal stabilization of the sensor.

Capacitive accelerometers utilize a spring-suspended proof mass, mounted between two capacitor plates. Any acceleration of the proof mass causes a small displacement, leading to a capacity change of the arrangement. The capacitive sensor is applied as frequency-defining component in an oscillating circuitry (with a typical bias frequency in the order of 10 kHz; used to reduce the low-frequency flicker noise contribution of the electronics). Even the slightest acceleration of the capacitive sensor causes a frequency modulation. The oscillator output signal is read out via a bridge arrangement with a LNA and demodulated to the base band. The minimum detectable acceleration of such an accelerometer is limited by the Brownian noise of the proof masses (and hence its temperature) and by the amplifier noise (Iafolla and Nozzoli, 2001). Thermal control of the sensor unit within at least 1 °C is required for such a system to allow for extremely low-noise performance.

10.23.4.22.1 Seismometer

Seismometers are specialized very low-noise accelerometers to detect seismic waves in planetary bodies. Seismic waves can be used to map the interior of planets and measure and locate quakes and other motions of the lithosphere (see also **Chapters 10.03** and **10.02**, the latter for internal structure modeling).

To define the propagation direction of a wave, usually, a three-axis seismometer is required. A servo principle is used: a proof mass, confined by electric forces, is kept at the equilibrium position by the control electronics in a closed loop arrangement. Any vibration of the proof mass, detected by capacitive, inductive, or optical means, is counteracted by the control electronics. Any change of the required electric force is recorded as the output signal of the seismometer.

For sophisticated seismic analysis, a seismometer network is required to constrain the origin, location, and propagation paths of the seismic waves, to disentangle surface waves from interior waves, to distinguish compression (P-waves) from shear waves (S-waves), and finally to constrain the interior structure of a planet or planetary body and its dynamic processes. Such a seismic network requires deployment from multiple planetary landers (cf. NetLander).

So far, only the Apollo seismic network has returned useful seismic information from the Moon (Lognonne, 2005). The Viking probes had simple seismometer mounted on their legs (Anderson et al., 1977), and the InSight mission foreseen for launch in 2016 (Banerdt et al., 2012) will bring a seismometer (and a heat flow probe) to the Mars surface (Mimoun et al., 2012). Miniaturized seismometers for planetary application are under development (e.g., Banerdt and Pike, 2001). Seismology as a tool for planetary exploration is discussed in **Chapter 10.03**.

10.23.4.23 Thermometer

Thermometers are used to measure the atmospheric temperature during entry and decent phase and at stationary landers and moving rovers. They can also be used to define *in situ* temperatures of planetary surfaces and are used to derive subsurface temperature and heat flow, for example, when integrated into a mole or drill system such as HP^3 (e.g., Spohn et al., 2014) Thermometers also play an important role as housekeeping sensors to monitor the temperature of instruments platforms or other spacecraft components. Many scientific instruments are sensitive to thermal variations, and hence, precise knowledge of the temperature is important to calibrate their measurement results.

Various types of thermometer principles are in use:

- Electric resistance thermometers
- Thermistors
- Thermocouples
- Microbolometers

The basic principle of resistance thermometer is that the conductivity of most conductors and semiconductors strongly depends on their temperature (in most cases a nonlinear dependence). By applying a small constant current, the resulting voltage over the resistor can be measured and with proper calibration the temperature derived. Such a sensor principle has been used, for example, to derive the precise atmospheric temperature during the decent of the Huygens probe in the atmosphere of Titan (Fulchignoni et al., 2005).

Thermocouples are used to measure temperature differences based on the thermoelectric or Seebeck effect. They can be used for wide temperature measurements but are rather limited in accuracy.

A microbolometer is a thermal sensor with very high resolution and sensitivity to IR radiation, hence a remote-sensing thermal sensor. The microbolometer reacts with a resistance change when exposed to IR radiation. This resistance change is measured and processed into temperatures. Microbolometer can be made very small and arranged in an array configuration allowing for thermal imaging – even without cooling of the detector array.

10.23.4.24 Pressure Sensor

Pressure sensors are used in atmospheres and exospheres to measure *in situ* the static pressure profile, derive the density profile during atmospheric descent, and measure the diurnal and seasonal changes of the ambient pressure on static or mobile landers and rovers. Two major sensor principles are in use:

- Semiconductor strain gage
- Capacitive systems

The semiconductor strain gages are thin Si-based membranes, which change their resistance due to deformation of the membrane caused by pressure changes. The piezoresistive changes of the detector are sensed in a bridge configuration with reference resistors to eliminate thermal drift. The very small bridge output voltage (e.g., 100 mV/350 mbar) is proportional to the sensed pressure and has to be amplified with a low-noise differential instrumentation amplifier.

Capacitive systems are based on the distance change of thin capacitor membranes (similar principles as a condenser

microphone), which results in a capacitance change. This change is usually sensed in an oscillator circuit where the sensor capacitor is used as a frequency-defining component. Hence, any change in pressure results in a change of the output frequency, which can be easily and precisely measured with a digital counter and precision timing gate.

The pressure sensors are used in three different measurement configurations:

- Absolute pressure
- Differential pressure
- Gauge pressure

For absolute pressure measurement, the sensor membrane is working against a reference pressure (e.g., vacuum) in a closed system. When the reference pressure is 1 atm, then the sensor is called a gauge pressure detector. For differential measurements, the membrane works in an open system and hence is sensitive to the pressure gradient.

The pressure profile instrument on board the Huygens probe (Harri et al., 1998) is an example of an absolute pressure capacitive system.

10.23.4.25 Microphones

Microphones are used for atmospheric acoustics (Petculescu et al., 2006) to study acoustic waves and acoustic events and to derive speed of sound. The detector used as a microphone in the Huygens permittivity and waves analyzer (Hamelin et al., 2000) is a differential pressure system based on a semiconductor strain gauge detector, which operates down to 90 K (Falkner 1999).

10.23.4.26 Microgas Sensors

Microgas sensors are foreseen to be used for *in situ* measurement on aerobots, landers, and rovers to detect and analyze specific constituents and trace gases of planetary atmospheres. Various principles for microgas sensors exist. Some are based on miniaturized IR sensors with emitters employing micromachined Fabry–Pérot interferometers, tuned to wavelengths of an absorption line of a certain gas. To resolve narrow absorption lines, a narrowband emitter coupled with a narrowband detector is required. The attenuation between the emitted signal and the received signal is measured to conclude on the concentration of the absorbing gas.

To achieve a narrowband emission, a microcavity with two mirrors (made, e.g., from CdHgTe) with a typical thickness of about 3–5 μm is employed. The microcavity is pumped with a laser diode (classically in the range from 0.6 to 0.9 μm wavelength), resulting in a beam with wavelength of 3–6 μm. As detector microbolometer can be used.

Another development is based on the so-called electronic nose principle, where integrated sensor arrays are employed. Individual sensor elements like polymer film are tuned to a specific gas component and respond with a change of volume and resistance when a specific gas reached the sensor. The resistance change can be easily measured. Using an array of different modified polymers creates very specific resistance change patterns, which are used to identify the gas. Electronic neuronal networks can be used to train the sensor to easier recognize a gas or gas mixture.

Microgas sensors and electronic noses are miniaturized alternatives to (rather bulky) gas chromatograph–mass spectrometer and currently under development for space applications and already widely used in laboratory and commercial applications.

10.23.4.27 Surface Dating

Luminescence dating is used to estimate the age of sediments by determining how long ago mineral grains were last exposed to daylight. The amount of unstable trapped electrons in the crystal structure of minerals like quartz or feldspar is used to define the age of the sediment. Trapped charge accumulates in minerals through time due to energy absorption from ionizing radiation like cosmic rays. Through heating (thermoluminescence dating) or optical stimulation in the case of the optical stimulated luminescence, the trapped electrons recombine and the energy is released by luminescence. The brightness of the luminescence corresponds to the amount of trapped charge, which is a measure of the total ionizing radiation received by the crystal since its last exposure to light (Aitken, 1998). With the knowledge of the ambient radiation dose rate (based on either measurement or modeling), the age can be derived.

The instrument consists of a heat or light source for stimulation and a photon detector with appropriate filter to detect the luminescence. Access to buried material and sample preparation are difficult to realize for robotic planetary missions. OLS instruments are widely used in geosciences on Earth, although spaceborne applications for use on Mars are under development (Kalchgruber et al., 2005).

Another approach to surface dating is based on crater counting as a technique for estimating surface ages over a more extended age range (see **Chapter 10.06**).

On Earth, the radiometric dating technique is widely used for surface dating. The same technique has been used also for Moon samples, which have been returned by Apollo and Soviet Luna missions. Unfortunately, this technique is not (yet) feasible for robotic space missions.

10.23.5 Instrument Suites for Exploration

Individual instruments as described in the previous section can never alone satisfy all objectives of a planetary exploration mission. In most cases, a combination of instruments is needed to provide new results when exploring our solar system planets and their moons and small bodies. All missions so far have had a more or less complex set of instruments on board. This approach will remain for future exploration missions, simply because most of the instruments need context information and additional data provided by other instruments on board the spacecraft.

In essence, all (geo-)scientific investigations performed on Earth would also be of interest on planets, moons, asteroids, and comets, but restriction due to technology and resource limitations (mass, power, volume, data rate, etc.) have limited what has been achieved up to now and also what can be done in the near future. Due to the high delta-v requirements of

planetary missions (see Chapter 10.17), the propellant consumes a large fraction of the overall spacecraft mass, often exceeding 50% or even 60%. Ever shrinking financial space-science budgets limit the size of future explorative missions to small and medium size. Giant missions such as Cassini/-Huygens are in the current budgetary environment unlikely for most of the world's space agencies. Selective and carefully considered priorities of the investigations are required. Focused science objectives and use of appropriate spacecraft and payload technology, paired with cautious trade-offs on mission risk and new technology, are desirable.

Smaller spacecraft require optimization of the spacecraft subsystems and in particular of interfaces between spacecraft and payload. Overheads on interfaces, implemented to simplify spacecraft engineering, integration, and assembly, compete with additional instrumentation and hence with the scientific performance of a mission.

An interdisciplinary view comprising spacecraft engineering, payload engineering, and science needs to be applied. A resource-saving approach is given by instrument suites tailored to the scientific requirements rather than collections of individual instruments (Figure 5). A careful instrument suite design based on well-defined scientific requirements allows for significant resource reduction and returns high-level scientific results even on smaller exploration missions.

The following section describes examples of instrument combination for various disciplines. The examples provided are intended as illustration and are under no circumstance exhaustive.

10.23.5.1 Planetary Size, Mass, Shape, and Interior

The precise measurement of a planet's size, its mass, and its shape, in particular deviations from the spherical shape by determination of the J-terms (zonal coefficients), and localization of mass irregular concentrations (mascons) are fundamental and also required for many other scientific investigations, as well as for precise spacecraft navigation. Interior structure and dynamics, in particular of the terrestrial planets (see Chapter 10.02), division of core/mantle/crust, and their motion have to be obtained. Although from gravity and magnetic measurement, some key information can be derived, seismic data measured on the surface are also important (see Chapter 10.03). Of course, landing on a planet is necessary to deploy a seismic station or even preferably a network of stations (see Chapter 10.03). Deployment of an instrument suite by means of a high-speed penetrator, as a cost-reducing alternative to soft landing, has been considered (e.g., Deep Space 2 Mars penetrator or JAXA's Selene – a lunar penetrator), but so far not successfully demonstrated in space. Small penetrators are under development (Vijendran et al., 2010) and were proposed for Juice mission or could be used for Mars missions as piggyback.

Measurement of the heat flow is important to constrain internal heating and temperatures. Temperatures have to be measured under the surface at various depths with high precision. The development of self-penetrating mole equipped with instrumentation to measure the heat flow (the heat flow and physical properties package HP^3) is ongoing (Spohn et al., 2014) and will be implemented in the InSight mission (Banerdt et al., 2012) to be launched to Mars in 2016. Magnetometers play an important role by measuring the magnetic field and hence electromagnetic material properties of the interior and constraining the internal field generation. Both measurements from orbit and at the surface are of importance.

10.23.5.2 Surface and Subsurface

Surfaces are analyzed via remote-sensing and *in situ* observations. A long list of interesting phenomena exists, resulting in a huge variety of possible instrumentation. Nevertheless, a kind of standard set of instrumentation has emerged, including imaging and observation in the electromagnetic spectrum (spectroscopy and radiometric measurements),

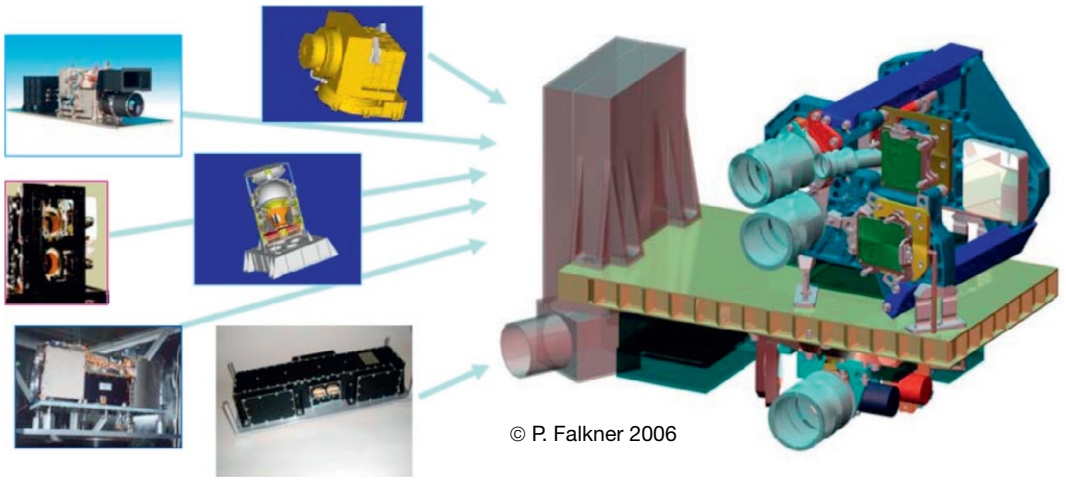


Figure 5 Evolution of collection of instruments into an optimized highly integrated instrument suite (HIPS). Stereo imager (middle top), laser altimeter transmitter (small optics, upper right), highly integrated broadband radiometer and imaging spectrometer (right bottom), and UV spectrometer (left bottom and left top) and supporting DPU with power supply (left bottom box between UVS and UVS).

from radio waves to γ-rays, mainly to define morphology and chemical composition. Laser and radar altimeters are used for topographic studies on the surfaces and to measure gravity tides and surface roughness (Table 2). GPR (cf. Mars Advanced Radar for Subsurface and Ionospheric Sounding (MARSIS) on Mars Express; Picardi et al., 2004) or SHARAD on MRO (Seu et al., 2007) is used for subsurface investigations because of the deep-penetration capabilities of low-frequency waves. An interesting application for GPR is the search for subsurface water or to investigate the ice–water interface like on Europa, although penetration of the full thickness of the ice layer is rather challenging. High-resolution imaging is important to provide photogeologic details and morphological and topographic information (particularly in combination with stereo imaging and laser or radar topography) and is also used for surface dating based on crater counting.

High-resolution spectral analysis in UV and NIR is used to define the chemical mineralogy possibly enhanced with elemental compositional information provided by XRF and GRS. Important factors are a high spatial resolution and the possibility of correlation with images provided by high- and medium-resolution imagers in the visual range, under the restriction that no dense atmosphere prevents optical access to the surface from an orbiting spacecraft. Imaging spectroscopy in the various bands with coaligned pointing to the visual imager and the star trackers, which are needed to derive the absolute pointing knowledge, can be best realized in an integrated suite approach. The instrument optics and focal planes are mounted very close together on common small platforms (Falkner et al., 2004; Kraft et al., 2005). This avoids structural problems (bending and thermoelastic expansion), which are more difficult to handle when distributed instruments and navigation subsystems are accommodated on an extended instrument platform. A simplified exchange of data from the instrument suite can be easily introduced in the design of the highly integrated payload suite (HIPS) data handling system.

Magnetometers are used to map magnetospheres and analyze their dynamics and measure fluctuations in the ambient magnetic field. Most recent developments of a single-chip low-power ASIC for complete three-axis FGM, using higher-order sigma-delta conversion and digital filtering (Magnes et al., 2008), demonstrate the possible route for miniaturization of instruments.

Another class of measurements can be done *in situ* on planetary surfaces. Surface access to planets remains challenging due to local environmental conditions (e.g., high temperatures on Venus or Mercury and radiation in the Jovian system), high delta-v requirements (e.g., Mercury, Galilean moons, and Titan), and overall technical complexity. Nevertheless, many missions have achieved successful landings and new ones are under preparation. Initial stationary landers such as the Viking landers on Mars (Soffen and Snyder, 1976) have been enhanced with mobile rovers (e.g., Mars and Moon) and self-penetrating moles, and drills are under preparation (Richter et al., 2005). Subsurface access has been just recently demonstrated by the Curiosity (MSL) rotary-percussive sample acquisition device. New instrumentation with significant miniaturization and the use of advanced technology, particularly state-of-the-art microelectronics, are required. Table 3 lists some of the instrumentation already used, planned, or under development for *in situ* observations.

Table 2 Instrument examples to constrain planetary size, mass, shape, and interior

Remote sensing	In situ
Laser or radar altimeter	Seismometer
Radio science and accelerometer	Heat flow experiment
Ground-penetrating radar	Magnetometer
Imager	

10.23.5.2.1 Surface station

A surface station is a stationary platform deployed on a planetary body, providing the infrastructure (power, thermal environment, data handling, communication, etc.) for the scientific instrumentation (see Table 3 under *in situ*). Deployment mechanisms are required to permit the instruments unobstructed access to the surface and give some degree of freedom for the sample selection and positioning. Several instruments require

Table 3 Instrument examples to analyze planetary surfaces and subsurface by means of remote-sensing and *in situ* instruments

Remote sensing	In situ
• High-resolution imager (visual)	• Panoramic camera
• Stereo camera	• Laser mass spectrometer
• Spectral Imager (UV, NIR, FIR, visual)	• LIBS/Raman spectrometer
• Surface and subsurface radar	• Alpha particle x-ray spectrometer
• Laser altimeter	• x-Ray diffractometer
• x-Ray, gamma-ray, and neutron spectrometers	• Attenuated total reflection spectrometer
• Microwave sounder, submillimeter sounder	• Mössbauer spectrometer
• Magnetometer	• IR Fourier spectrometer
• Thermal infrared (IR) radiometer	• Gas chromatograph and mass spectrometer
	• UV spectrometer
	Microscopic imager
	• Heat flow and physical properties package
	• (Miniature) Thermal emission spectrometer
	• Magnetometer
	• Seismometer
	• Ground-penetrating radar (surface deployed)

very specific access to and alignment with the sample (e.g., LMS, APX, and Mössbauer) and hence a precise deployment mechanism. Imager and panoramic imager should be mounted on a deployable boom to provide an elevated view with a mechanism providing elevation and azimuth rotation, required for panoramic imaging. A thermometer, wind meter (anemometer), magnetometer, and many other instruments require some distance from the surface station, simply to reduce the influence and disturbances on their measurements, and are therefore preferable mounted on a (deployable) boom. A robotic arm can introduce a wider range for positioning of instruments (cf. Beagle II; Maki et al., 2011; Pullan et al., 2004).

A geophysical surface station is going to be launched to Mars in 2016. InSight (Banerdt et al., 2012) will carry a seismometer (Mimoun et al., 2012) and a heat flow package (Spohn et al., 2014). This basic payload is augmented by a radio-science package, a magnetometer, and a pressure sensor. Geophysical network stations have been discussed at NASA, ESA, CNES, and JAXA for many years (e.g., Mesur, MarsNet, Intermarsnet, and NetLander) but have never been flown. InSight will provide the first dedicated geophysical lander and may pave the way for a lander network in the future.

10.23.5.2.2 Rover

An extended range of mobility can be achieved with a rover. Impressive examples are given by both MER rovers and recently MSL Curiosity. Free-traveling rovers require their own local very efficient power supply (batteries, RTG, or solar panel), which provides energy for roving and all the required subsystems and to allow for thermal management. A remote link via a relay satellite or direct-to-Earth is required for data transfer and commanding and sufficient on-board autonomy for navigation and traveling are mandatory, as usually only limited communication windows with ground control exist. A surface station is a stationary platform deployed on a planetary body, providing the infrastructure (power, thermal environment, data handling, communication, etc.) for the scientific instrumentation (see **Table 3** under *in situ*). Deployment mechanisms are required to permit the instruments unobstructed access to the surface and give some degree of freedom for the sample selection and positioning. Several instruments require very specific access to and alignment with the sample (e.g., LMS, APX, and Mössbauer) and hence a precise deployment mechanism. Imager and panoramic imager should be mounted on a deployable boom to provide an elevated view with a mechanism providing elevation and azimuth rotation, required for panoramic imaging. A thermometer, wind meter (anemometer), magnetometer, and many other instruments require some distance from the surface station, simply to reduce the influence and disturbances on their measurements, and are therefore preferable mounted on a (deployable) boom. A robotic arm can introduce a wider range for positioning of instruments (cf. Beagle II; Maki et al., 2011; Pullan et al., 2004); navigation and traveling is needed, as usually only limited communication windows with ground control exist. The locomotion system with the required structures and the collection of all needed subsystems drive the minimum size and mass of such a rover (cf. MER, ExoMars, MSL) in a range from 150 to 1000 kg, allowing for an accommodation of an enhanced and meaningful payload suite.

A small rover can carry only a much miniaturized payload suite (e.g., APX and Mössbauer and a microscopic camera). In the simplest form, the rover is connected to the surface station via a thin tether for power supply, data transfer, and commanding. Due to the tether, the range is limited to several tens of meters (cf. Nanokhod; Bertrand et al., 2000). Small untethered rovers are limited to the near surroundings of the landing platform (cf. Sojourner rover) and have limited obstacle negotiation capabilities.

10.23.5.2.3 Subsurface access

Access to the subsurface is of interest to understand layering, have access to unaltered material, measure heat flux from the interior of the body, and define soil characteristics. Nonintrusive access is given by means of subsurface-penetrating radar or to some limited extent also with a microwave sounder. A drill allows for local subsurface access either to deploy embedded sensors or to retrieve samples (cf. ExoMars drill and MSL drill; Maki et al. (2011)). Another possibility is the use of a mole (cf. Pluto on Beagle 2). On ice-covered bodies (e.g., comets, Europa, and Mars poles), a melting probe can be used for ground penetration by sampling the medium as the probe melts through the ice (Kaufmann et al., 2009).

10.23.5.2.4 Mechanical tools

Mechanical tools are used for removal of the top layers on samples to provide access to unaltered pristine material and are often referred to as a 'geologist's hammer.' A grinder is a rotating device able to create a hole of a few centimeters in diameter and a few millimeters in depth in rocks, which is then further explored with *in situ* instruments (cf. Rock Abrasion Tool (RAT) on Mars Exploration Rover (MER) or the corer-grinder on Beagle 2; Richter et al. (2002)). Understanding the differences between the original rock surface and the interior reveals information about the rock formation process and environmental conditions, which have altered the outer layer of the rock.

Another concept is the ultrasonic drill with a piezoelectric-driven mini hammer tip, which pulverizes the material for further analysis (Chipera et al., 2003).

10.23.5.3 Atmospheres, Exospheres, and Ionospheres

To understand the complex dynamics of atmospheres and their relationship with the surface, an extended set of instrumentation is essential. Also, for atmosphere research, the combination of remote-sensing and *in situ* observations is applied. Balloons, commonly used in the Earth's atmosphere for *in situ* sounding, are yet rarely used in planetary atmospheres as instrument platforms (e.g., VEGA 1 and 2 at Venus 1985). Entry and descent probes (e.g., Galileo at Jupiter, Huygens at Titan, and various Venus probes) have delivered significant data due to their capabilities to provide *in situ* compositional and thermodynamic profiles. Lander instruments play a key role in atmospheric measurements because they offer unambiguous observations of diurnal and meteorologic variations, in particular for surface–atmosphere interactions, which are difficult or impossible to observe from orbiting spacecraft by means of remote sensing (Morgan et al., 1995). Many open questions remain, due to the dynamic changes associated with

atmospheres, and many observations over long periods of time are mandatory, while comparative interpretations on multiple planets are required. Networks of surface stations capable of long observations, long-duration aerobots cruising in various altitudes, and the deployment of microprobe swarms are important for the next steps in exploration. The key problems with very small stations and microprobes are the maintenance of thermal conditions within hostile environment, the energy supply, and communication (in particular energy requirements for communication and data transfer) with the data relay (direct-to-Earth communication is difficult due to the long-range and associated space loss, the required RF power, and the energy density limits of current energy storage systems). Some key technologies for microprobes are under development such as localization and communication (Wells et al., 2004).

Remote sensing of atmospheres is performed by nadir observations and also limb sounding. **Table 4** lists a few key instruments for remote-sensing and for *in situ* investigations. Imagers are important to observe the dynamics of an atmosphere. Very high resolution is usually not required; a wide FoV is preferred to guarantee longer-term observations of time-dependent variations and dynamic processes. Spectrometers are used to study the composition in various wavelengths. Atmospheric weighting functions have to be applied to perform sounding at different altitudes. These functions are describing the layer of the atmosphere, from which the radiation measured at a particular wavelength was emitted (Elachi, 1987, pp. 302–311). Measurements performed at various wavelengths reflect peak emissions from different altitudes. Similar principles are used for microwave and submillimeter sounders. Microwave sounders are in particular sensitive to reduced forms of oxygen and nitrogen (water and ammonia). Laser altimeter can be used to define the local atmospheric opacity and atmospheric aerosol loading (Ivanov and Muhleman, 1998).

Radar signal transmission through an atmosphere is dependent on the TEC with additional phase shifts and damping occurring in the ionospheres. Both effects are exploited with radar measurement. Precipitation and water vapor content measurements are also classical applications for radar observations. Also, limb sounding and observations by occultation (e.g., of telecommunication radio link signals) are used by radio science to define, for example, the density profile of an atmosphere.

Lightning detectors are accommodated either on orbiting spacecraft or flyby spacecraft (e.g., Radio and Plasma Wave Science on the Cassini orbiter; Fischer et al., 2005) or by *in situ* instrument such as the HASI permittivity, waves, and altimetry analyzer (PWA), installed on the Huygens probe (Fulchignoni et al., 2005).

A most recent example of a mission dedicated to study the atmosphere and its interaction with the solar wind is Venus Express. Juno is another example containing significant instrumentation for atmospheric sciences to investigate Jupiter (Bolton and The Juno Team, 2012).

10.23.5.3.1 Entry and descent probes

Every *in situ* access to a dense atmosphere and to the surface of a planet with a dense atmosphere requires an entry and descent vehicle. The entry vehicle has to suffer a significant heat flux when entering the atmospheric interface due to the high entry velocities (e.g., around 9.8 km s^{-1} for Venus entry or minimum of 47 km s^{-1} for Jupiter), when performing a hyperbolic insertion. Entry is very challenging in particular for the large gaseous planets with their strong gravity field. A robust ablating thermal protection system is required to keep the subsystems and instruments within an acceptable temperature range during the entry. In some cases, the overall thermal protection system with required structure exceeds even more than half of the overall system mass of the entry vehicle. The remaining mass must be shared by all required subsystems (power, data handling and communications, etc.), leaving only a small fraction for scientific instrumentation. Miniaturization is a key prerequisite to be able to accommodate a meaningful payload also on smaller systems. Access to the environment is limited during the entry phase as the probe is usually cocooned and thermally fully isolated. Acceleration can, however, be measured in this cocooned state to define the drag in order to derive the ambient atmospheric density. Thermal sensors and strain gages embedded in the ablative material are employed to understand engineering details like the ablation processes during the entry. After the deployment (jettison) of the aft and front heat shield, all scientific instruments can be deployed. As soon as the RF system is operable (in particular after the normal blackout period), important data on the entry trajectory can be obtained by exploiting the Doppler shift of the RF-carrier signals. An example for an entry sequence and consecutive deployment of scientific sensors is given by the Huygens probe, which successfully entered the Titan atmosphere

Table 4 Instrument examples to investigate planetary atmospheres, exospheres, and ionospheres via remote-sensing or *in situ* analysis

Remote sensing	In situ
• Wide field imager (visual)	• Aerosol and dust analyzer
• Spectrometer (UV, NIR, FIR, visual)	• Pyrolyzer and mass analyzer
• Laser altimeter	• Magnetometer
• Microwave sounder, submillimeter sounder	• Nephelometer, miniature gas sensors and electronic noses
• Radar including ionospheric sounder	• Miniature gas sensors and electronic noses
• Radio science	• Anemometer
• Lightning detector	• Up- and down-dwelling fluxes
	• Composition (gas chromatograph and mass spectrometer)
	• Pressure, temperature, acceleration sensors
	• Space plasma and energetic atom analyzer

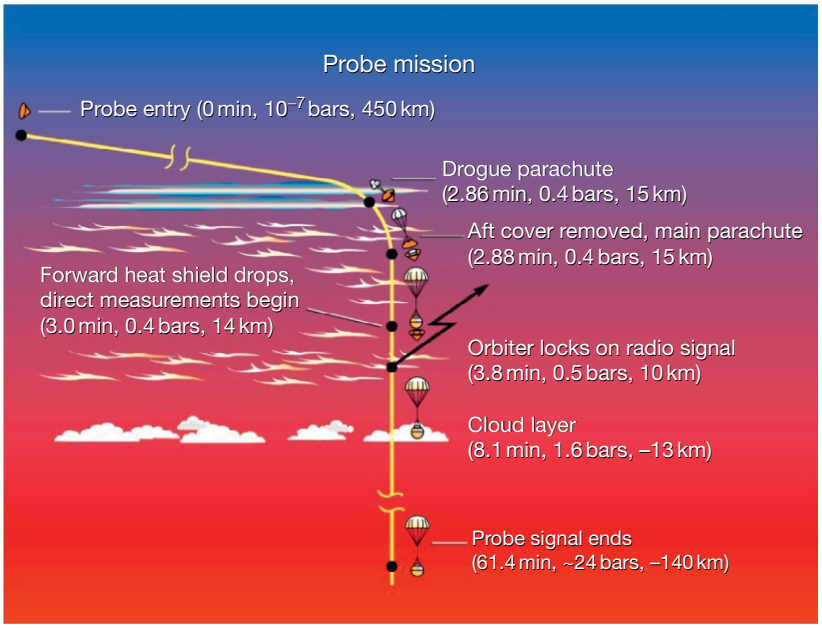

Probe mission
Probe entry (0 min, 10^{-7} bars, 450 km)
Drogue parachute
(2.86 min, 0.4 bars, 15 km)
Aft cover removed, main parachute
(2.88 min, 0.4 bars, 15 km)
Forward heat shield drops,
direct measurements begin
(3.0 min, 0.4 bars, 14 km)
Orbiter locks on radio signal
(3.8 min, 0.5 bars, 10 km)
Cloud layer
(8.1 min, 1.6 bars, –13 km)
Probe signal ends
(61.4 min, ~24 bars, –140 km)

Figure 6 The Galileo probe mission events diagram. Image NASA-JPL.

on 14 January 2005, or by the Galileo entry probe (Figure 6). Both the Huygens probe and the Galileo probe had been designed as descent probes, able to deliver science data during the descent in the atmosphere of Titan and Jupiter, respectively.

The following list summarizes typical instruments used on such kind of descent probes (Table 5).

In many cases, the entry probes are optimized to deliver a lander (cf. Beagle), rover (cf. MER, MSL, ExoMars), or aerobot, with a strongly reduced or even no scientific payload at all installed on the probe itself, but only a dedicated payload on the delivered unit.

Table 5 Typical payload examples for entry and descent probes in planetary atmospheres

Entry and descent probe payloads
• Atmospheric structure package (accelerometer, gyros, pressure and temperature sensor, recession measurement)
• Miniaturized (spectral) imager
• (Polarization) Nephelometer
• Anemometer
• Up- and down-dwelling fluxes
• Miniaturized gas chromatograph and mass spectrometer
• Aerosol analysis
• Energetic particles and plasma spectrometer
• Doppler wind experiment (via radio science)
• Acoustic instrument (e.g., to define speed of sound)

10.23.5.3.2 Aerobots

Aerobots are aerial robotic platforms (e.g., Vega balloons) providing an excellent platform for *in situ* sensing instruments at specific altitudes in the atmosphere of planetary bodies. Aerobots have to be delivered by an entry probe before they can be deployed at the target altitude in the atmosphere of interest. Various aerobot principles exist although not many have been flown in space so far. The lighter-than-air balloons (light gas balloons, airships, blimps, and Montgolfier) are based on the buoyancy effect using a gas or gas mixture, which is lighter than the surrounding atmosphere. Heavier-than-air vehicles (gliders, rotating wings, and airplanes) are based on a relative motion through the atmosphere to provide a lift. The main challenges for both principles are the deployment of the balloon after entry and the survival of the local environmental conditions over the required observation time.

Almost all of the *in situ* instruments listed in Table 4 can also be used on an aerobot in a miniaturized form to measure, for example, atmospheric composition, horizontal and vertical winds, layering, cells fluxes, density, pressure profiles, gas mixture, elemental sources and sinks, and eventually atmospheric escape processes. Certainly, it is also possible to perform remote observations of the surface, provided the atmosphere is not completely opaque. Payload miniaturization is required as the available payload mass fraction on an aerobot is rather small. The aerobot itself is limited in (undeployed) size and mass to be accommodated within the entry probe, which on its own is already significantly mass-constrained due to the high mass fraction of the heat shield and thermal protection system (e.g., Chassefière and The EVE Team, 2009). Aerobots are also a possible platform for deployment of a single or a swarm of microprobes.

10.23.5.4 Magnetosphere

Investigation of the planetary magnetism is about characterization of frozen ambient fields in bodies, which have cooled well below the Curie point and the magnetic fields originating from a planetary dynamo (see **Chapter 10.07**). Within the planetary magnetospheres, the magnetic field dominates the behavior of electrically charged particles (Kivelson, 1981). Interesting on

one side is the interaction of the planetary fields with the solar wind and its magnetic field, resulting in a bow shock followed by the magnetopause on the upwind side and the long turbulent magnetotail on the downwind side, the magnetically trapped particles in the radiation belts, direct entry of solar wind particles into the magnetosphere through the polar cusp (aural effects), and interaction of the field with any associated moons. On the other hand, the magnetic field provides significant data about the (deep) interior of the planet.

Instrumentation follows these science lines and concentrates on the definition of the internal magnetic field to constrain the interior of the planet or to investigate the magnetosphere and the magnetohydrodynamic (MHD) processes, plasma and waves, reconnection, substorms, and turbulences in it.

For investigation of the internal field, a combination of spacecraft at different altitudes is preferred: one spacecraft on a circular polar orbit to measure precisely the internal field, a second spacecraft on an elliptical polar orbit to understand the interaction with external fields, and possibly a surface station, which provides *in situ* reference of the local surface field.

To investigate magnetospheric and plasma processes, a tetrahedron formation with four similar spacecraft with identical instrumentation is of use, such as ESA's Cluster II mission (Escoubet et al., 2001). Spinning spacecraft is preferred to accommodate long wire booms, which derive their attitude stabilization by the gyroscopic stiffness. The spin also provides 360° scan for all instruments with a radial bore sight without moving mechanism.

Space plasma dynamics involve processes on four scales: (1) the global scale, (2) the MHD scale, (3) the ion scale, and (4) the electron scale. Observation of processes can be best achieved with four spacecraft in a tetrahedron formation like Cluster II, but at least at three scales (MHD, ion, and electron scales) at the same time, thus requiring up to 12 spacecraft in total (cf. MMS Mission, Scope, or Cross-Scale). Instrumentation on a particular scale can be identical, but differ between the scales to cope with particular particle types, energy ranges, and dynamics associated with the processes at the different scales (Table 6).

10.23.5.5 Moons

Payloads for moons are basically similar to what has been described already for their mother planets, although additional topics such as the interaction of the moon(s) with the mother planet (and their rings), generated tides and associated tidal heating (cf. Io and Europa), gravity perturbations or mutual influence and interaction with magnetosphere (cf. Ganymede), and particles (cf. Io; Spencer et al., 2002) are of particular additional interest. As an exception, Titan is the only moon in our solar system with a dense atmosphere, and hence, entry probes with particular tailored payloads (cf. Huygens; Lebreton and Matson, 2002) are possible (Table 5).

Due to the lack of a dense atmosphere around the other moons, landing requires full propulsive cancellation of the orbital and gravitational velocity with chemical motors. Despite the relative lower gravity of most of the moons, high propellant mass has to be carried on board the spacecraft because of the transfer, insertion, and landing delta-v requirements, leaving only a relatively small mass ratio available to the surface payload.

Table 6 Examples of instrumentation for magnetospheres and plasma processes.

In situ
• Magnetometer (AC and DC field)
• Plasma particle package
• Electron drift instrument
• Active spacecraft potential control
• Plasma wave analyzer
• Neutral and ionized particle analyzer
• Radiation monitor
• Fast and hot plasma analyzer

10.23.5.6 Rings

As we know by now, each of the giant planets in our solar system possesses a more or less complex and unique ring system (Figure 7). Composition, origin and age, characterization of the physical properties, in-ring collision and collision processes with external material (meteorites), evolutionary processes, particle size, and distribution are some of the key topics to be addressed in the further understanding of ring systems (Gordon et al., 2002).

Ring observations usually have to be shared or compete on board the spacecraft with other instrumentation but can be performed also with Earth-based and space telescope observations (de Pater et al., 1999). Most of the instruments listed in Table 3 (under remote sensing) can in principle be used for ring observations, of the shepherd satellites and their gravitational interaction with the rings and the central body. Imaging, and in particular close-up imaging, and compositional analysis with UV and IR (imaging) spectrometer play a key role. Long-term observations, required for the understanding of ring dynamics, can be mainly sustained by observation from Earth. *In situ* observations can be done with a dust analyzer, as for the first time performed on Galileo, providing dust spatial densities, grain size distribution, and impact speed along the spacecraft trajectory (Krüger et al., 1999).

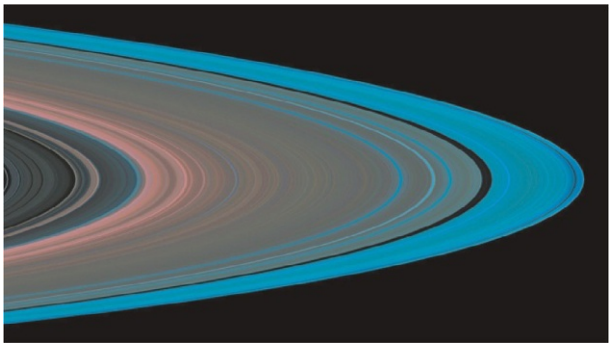


Figure 7 Saturn's ring system observed in the UV by Cassini's ultraviolet imaging spectrograph. The turquoise represents more water ice in the rings. Image NASA/JPL/University of Colorado 'PIA05075: Saturn's A Ring from the Inside Out.'

10.23.5.7 Asteroids and Comets

Comets represent the most fundamental building blocks of the solar system preserved in a low-temperature and low-gravity environment (e.g., Schulz, 2002). Most of them remain in the Oort cloud or in the Kuiper Belt outside the orbit of Neptune (see also **Chapter 10.16**). Only occasionally an object is randomly scattered into the inner solar system by gravitational perturbations, becomes active and bright, and can therefore be observed with telescopes from Earth and space or visited by spacecraft (cf. Giotto, Deep Space 1, Stardust, Deep Impact, and Rosetta). The outer comets can hardly be visited with close encounters as the travel time gets too long and the high escape velocity of the spacecraft necessary for reaching the outer parts of the solar system would require high delta-v maneuvers for transfers to a comet (cf. **Chapter 10.17**). This is hardly feasible with current propulsion technology. Mainly comets crossing the inner solar system on highly elliptical orbits can be visited. While flybys are relatively easier to achieve, orbiting missions, such as the Rosetta mission to comet 67P/Churyumov–Gerasimenko, still have high delta-v requirements because the spacecraft has to have a similar angular velocity as the comet in elliptical orbit around the Sun.

Asteroids have been divided into various classes depending on their distance to the Sun, or more specifically the type of orbit (Chapman, 2005). The asteroids in the main belt are believed to be remnant planetesimals with complex collisional and dynamic activity, whereas NEOs may be either asteroids or dormant comets (**Figure 8**). They come close to or can be on Earth-crossing orbits and are therefore potentially hazardous for a collision with Earth. Tracking and observations from Earth are important for determining the orbit ephemerides and gross physical characteristics of these bodies, such as size, shape, rotation period, and spectral class, which are important prerequisites for choosing the appropriate and most interesting targets for exploration missions.

In a flyby or orbiting mission to small solar system bodies, an accurate determination of their dimensions and rotational properties, as well as their surface, topography, and morphology, can be obtained by direct imaging in the visible band. The interior can be investigated with GPRs, like with the Comet Nucleus Sounding Experiment by Radiowave Transmission (CONSERT) experiment (Barbin et al., 1999), a bistatic radar system with the object between transmitter and receiving radar to achieve a tomography of the body.

A radio-science experiment is necessary to define the gravitational field and moments of inertia of the body and for absolute position definition, which is required in relation with altimetry measurement by radar or LIDAR. For the compositional analysis of the surface by remote sensing, multiwavelength spectroscopy is required for wavelengths from γ-rays to submillimeter range. These spectrometers can also be used to determine the outgassing rate of comets and search for activity around asteroids. Surface and subsurface temperatures can be defined with thermal IR and microwave spectrometers. The long list of possible *in situ* instruments can be used to provide local measurements on the surface and, in the case of active bodies, by flying through the coma and tail, determining the composition of the gas and dust particles ejected from the nucleus. By use of drills and moles, subsurface characterization can be achieved in a similar style as on planetary surfaces.

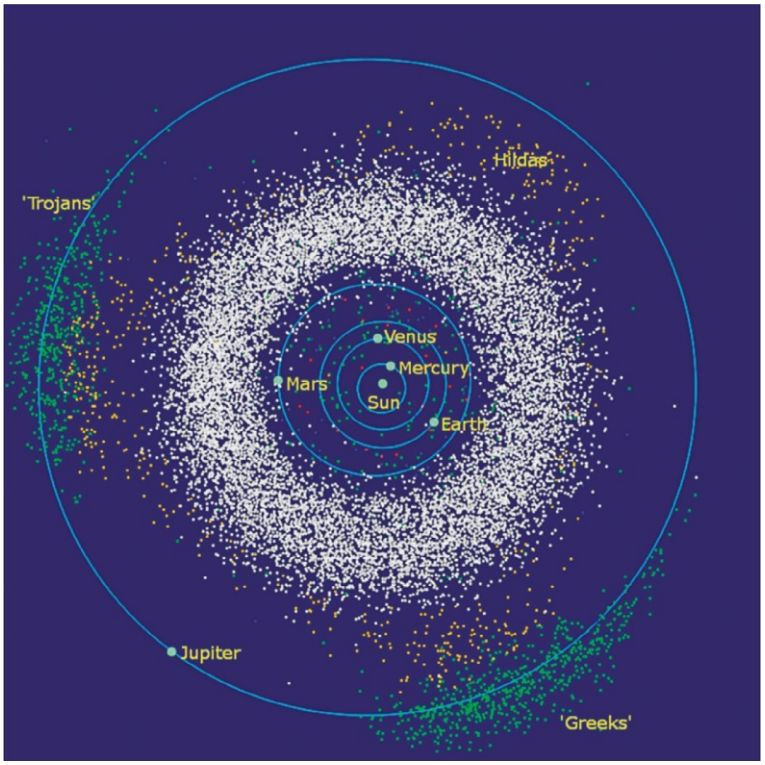


Figure 8 The main asteroid belt concentrates mainly between Mars and Jupiter. Image NASA-JPL.

Table 7 Instrument examples for remote or *in situ* sensing payloads for asteroids and comets

Remote sensing	In situ
• Imaging system • Nucleus sounding experiment by radio wave transmission • Mass spectrometers for compositional analysis of neutrals, ions, and dust particles • Radio-science experiment • UV, visible, and IR imaging spectrometers • γ-Ray, neutron, and x-ray spectrometers • Microwave sounder • Laser or radar altimeter • Dust particle analyzer	• Panoramic and microscopic imaging system • Ion mass analyzer • Impact analyzer and dust accumulator • Spectrometer for ion and neutral analysis • Ion composition analyzer • Ion, neutral, and electron spectrometer • Langmuir probe • Magnetometer • Conductivity and permittivity probe • Conductivity probe • Plasma instruments • Alpha x-ray spectrometer • Laser mass spectrometer • Evolved gas analyzer – elemental and molecular composition • Physical surface and subsurface properties • Atomic force microscope • Penetration probes and drills • Seismometer • Acoustic monitor

Table 7 summarizes potential instruments for missions to asteroids and comets.

10.23.5.8 Exobiology

The field of exo- or astrobiology (see Chapter 10.12) combines the key characteristics of astronomy, biology, and geology, and as such, so does the instrumentation. It not only comprises most of the instrumentation described so far but also involves laboratory activities, as well as in-field search of 'exotic' life forms on Earth (extremophiles), extending also to the search and characterization of extrasolar planets. Within our solar system, the focus resides on Mars, Europa, and Titan, due to the presence or assumed presence of water (cf. Europa) and organics (cf. Titan). Search and analysis of organics, search for water as prerequisite for life, search for habitable zones by remote sensing, and *in situ* search include most of the instruments given in Table 3 but require further improvement and new instrumentation, as summarized in Table 8. Due to the harsh conditions on most of the planetary surfaces, the search is focusing on subsurface niches (e.g., with GPR), requiring mobility and subsurface penetration capabilities (e.g., drills or moles), equipped with sampling mechanism and dedicated sensors for analysis of prebiotic chemistry, as well as for the search for past or present life.

Table 8 Instrument examples for exobiology payloads suites for remote or *in situ* sensing

Remote sensing	In situ
High-resolution imager (visual)	Panoramic camera
Stereo camera	Laser mass spectrometer
Spectral imager (UV, NIR, FIR, visual)	LIBS/Raman spectrometer
Subsurface radar	Alpha particle x-ray spectrometer
x-Ray, gamma-ray, and neutron spectrometer	x-Ray diffractometer
Microwave sounder, submm sounder	UV spectrometer
Magnetometer	Attenuated total reflection spectrometer
Thermal IR radiometer	Mössbauer spectrometer
	IR Fourier spectrometer
	Gas chromatograph and mass spectrometer
	Microscopic Imager
	Heat flow and physical properties package
	(Miniature) Thermal emission spectrometer
	Radon exhalation
	Magnetometer
	Seismometer
	Ground-penetrating radar (surface deployed)
	Drill system
	Life marker chip
	Radiation monitor
	Permittivity probe
	Environmental package

10.23.5.9 Sample Return

A sample return is a specific case, since the emphasis is on the return of material to be further analyzed in dedicated terrestrial laboratories, where the same strong resource restrictions are not given as for space mission instrumentation. Not only significant higher resolution in all measurements can be achieved but also

different techniques can be applied, which cannot (currently) be used on robotic missions (Sears et al., 2002). The benefits have to be traded with the increased mission complexity, as it requires a spacecraft performing the transfer to the body, soft landing on it (involving either atmospheric entry where a denser atmosphere exists or propulsive landing on atmosphereless bodies), a dedicate sample mechanism, and sample transfer unit, which brings the sample to a container, embedded in the Earth Return Vehicle (ERV). The ERV does the ascent and escape from the body (or planetary system) and return transfer to Earth, where the Earth return container has to perform the reentry and safe landing on Earth.

In a minimum configuration, there will be limited additional scientific instrumentation foreseen on board the spacecraft, for navigation purposes in particular in proximity to the body and landing and sample site selection, and some minimal instrumentation for local sample selection. Context information for where the sample is taken is also required. For the sampling process itself, the sample mechanism and its interaction with the local environment are of key importance. The easiest case is to take a sample from the atmosphere. More complex is the sampling process for the terrestrial planets, as well as moons, asteroids, and comets, requiring interaction with the surface material (e.g., regolith, ice, and rocks).

In a low-gravity environment, either an anchoring system or active propulsion is required to provide stability to the landed platform, able to counteract the introduced forces by the sampling process, or a touch-and-go procedure is applied. Of course, a low-gravity environment has the benefit of a simplified landing and return with small delta-v. In a high-gravity environment, an escape stage is needed to provide enough velocity to leave the body's gravitational field, or at least to return to an orbiter and perform a rendezvous maneuver for sample exchange.

10.23.5.10 Engineering Aspects and Operations

Usually, the payload is one of the main mission drivers for cost, mission size, and risk. It is therefore very important to understand and reflect the detailed science requirements and consider all the drivers of these requirements on the spacecraft. The ground resolution of the imaging payload, for example, might be the main driver for the spacecraft attitude control, and hence, a detailed trade-off between science requirement and implementation complexity must be performed in an iterative process. A system engineering approach with an end-to-end view is required to properly manage all interrelations and dependences, requiring interdisciplinary knowledge (Cruise et al., 1998).

10.23.5.10.1 Engineering constraints

Resources such as mass, power, data rate, data volume, and in particular cost strongly limit planetary missions. The selection of instruments must be performed in the first instance to satisfy the defined science requirements of a particular mission within the estimated available resources. The use of the latest technology might help obtain more science per resource, but is not necessarily at lowest risk and cost. A well-considered trade-off during instrument selection and instrument design is required.

The environmental conditions at the target of observation and also during transfer phases are a strong driver for the design and selection of equipment and technology for the complete spacecraft and the instruments in particular. Radiation, thermal ranges, shock, and vibration levels are just a few of the drivers that strongly influence the design solutions of instruments and instrument suites. The payload is commonly the most sensitive part of a spacecraft. Instruments require specific conditions during operation, standby, and nonoperation (thermal ranges, vibration levels, power supply stability, low-electromagnetic-cleanliness (EMC) noise environment, etc.). Different instruments require different environmental conditions, hence driving the complexity of the spacecraft and of its subsystems. In particular, the front-end instruments (optics and focal plane and sensors) have the most stringent requirements.

Coalignment of the FoV between different instruments and time synchronization of the measurements are additional drivers, requiring stable instrument platforms and precision timing capabilities.

Many low-noise sensors require either active cooling (e.g., Peltier element or sorption cooler) or passive cooling via a radiator, because noise increases with temperature. In particular, with respect to planets with a hot environment (Mercury and Venus), the thermal design becomes rapidly a strong mission driver.

Sensors requiring direct access to the environment (e.g., thermal sensor, plasma instruments, and magnetometers) must be in many cases mounted outside the protective spacecraft and are consequently exposed to harsher environmental conditions (radiation, temperature, dust, and plumes) and may require a boom for deployment. Long appendages such as booms or antennas (in particular for GPR) require mechanism for deployment and change, when deployed, the spacecraft attitude stability, and moment of inertia.

Due to long transfer times, aging, radiation, and changing environmental conditions, instruments are subject to performance alteration, drift, and increased noise levels. A calibration is required, and in many cases, regular calibration cycles with reference sources have to be performed during the full lifetime of the instrument. Stable calibration targets have to be used for reference, and repointing of the instrument to the targets requires in many cases additional mechanism.

10.23.5.10.2 Operational aspects

All the aforementioned points have to be considered not only in the design of the instrument and instrument suite but also during operation. The onboard instrument data handling system and control computer need to operate the instruments within the required conditions and have to avoid any hazardous conditions. This can be performed either with ground control and commanding or via computer-installed autonomy on board the spacecraft. Data processing is used to produce telemetry packets out of the raw sensor data, which can be transferred via the telemetry link back to Earth ground stations. In many cases, the telemetry link is not available permanently, but limited to so-called communication windows; hence, data must be stored on board the spacecraft to prevent a loss of important scientific data. A data management system is required, which takes care of the data acquisition, buffering, and transfer to the telemetry link, when available. Some

instruments require already-onboard context information from other instruments to adapt and optimize their own operation. A data exchange and data flow management must be installed in such cases.

The telemetry link is limited in data rate and data volume. Depending on the nature of the data, loss or lossless compression can be used to reduce the data volume of an instrument. For the final data analysis, it is important that no compression artifacts are introduced or that these artifacts are properly understood and acceptable.

Housekeeping data, permanently transferred together with science data, summarize the current environmental and status conditions of the instrument and the entire spacecraft, such as temperature, supply currents, voltages, deployment state, pointing elevation, and the current operational mode. This is all of utmost importance to understand later the remotely acquired data set and perform successful data calibration and interpretation.

10.23.5.10.3 Verification and testing

Not only because of the unique nature of a space mission and the high mission cost but also because most of the instruments and instrument suites are prototypes, a thorough verification (in particular against the environmental conditions) and testing after the assembly and integration are mandatory. An instrument has to be qualified at harder conditions and verified at least at similar conditions as expected during launch, transfer, and operation of the space mission. Thermal-vacuum testing, vibration testing, and radiation testing are just some of the tortures that a space instrument has to survive without damage and serious performance degradation. The particular test procedures and conditions depend finally on the nature of the mission and have to be planned and characterized during the development phase of the mission.

10.23.6 Outlook and Timeline

All our planets, as well as a number of moons, comets, and asteroids, have by now been visited by spacecraft. What remains is a more thorough exploration of all objects to further improve our knowledge and understanding. Landing on planetary surfaces is required to provide ground truth for already existing remotely sensed data (e.g., Mercury, Galilean and major Saturn moons, other asteroid classes, and comets) and provide *in situ* observations not possible for orbit. Long-term observations, deployment of meteorologic, and seismic networks on neighbor planets (Venus and Mars) should also help better understand our own planet and its environment.

Sample and return from the terrestrial planets, various moons, and selected asteroids are further goals to be achieved. Sample and return missions are definitely also technology precursor for long-term future human exploration of, for example, Mars or asteroids. The further exploration of the interstellar heliopause at some 200 AU from the Sun is another longer-term goal.

Due to the extended duration of space mission development, most of the shorter-term and medium-term goals are already under study, planning, or development and embedded in the various national and international space plans. The realization of these plans strongly depends of course on the political environment and the will and capabilities to financially support these plans.

Technology development is an important prerequisite to prepare further spectacular and striking space missions, which are able to add scientific value and allow further exploration. Research and development is definitely a longer-term endeavor, which needs financial support based on long-term planning and visions. Technology must be available or in reach when a mission planning starts, to avoid delays, fallback solutions, cost overruns, or in worst cases a mission failure.

Miniaturization and high levels of integration of instruments and spacecraft subsystems similar to what is ongoing in the consumer world; wider use of nanotechnology, systems on a chip, and MEMS; further development of space-qualified application-specific integrated circuits (ASIC) and FPGA, new detectors, sensors, and materials; and adaptation and qualification of terrestrial instrumentation principles for space application are important midterm steps in preparation of the future of planetary exploration.

Finally, it is our imagination, when coupled to the successful completion and promotion of current space missions, which will prepare the ground for future exploration.

References

Aitken MJ (1998) *An Introduction to Optical Dating. The Dating of Quaternary Sediments by the Use of Photon-Stimulated Luminescence*. London: Oxford University Press.

Albee AL and Bradley JG (1987) SEMPA – Scanning electron microscope and particle analyzer for the CRAF mission. *Lunar and Planetary Science* XVIII: 13–14.

Ameron FHW, Chaudhary A, and Short RT (2012) Low-cost micro mass spectrometers for handheld chemical analysis and distributed networks for space flight missions. In: *International Workshop on Instrumentation for Planetary Missions*, abstract; http://www.lpi.usra.edu/meetings/ipm2012/pdf/1040.pdf.

Anderson DL, Miller WF, Latham GV, et al. (1977) Seismology on Mars. *Journal of Geophysical Research* 82: 4524–4546.

Balogh A, Beek TJ, Forsyth RJ, et al. (1992) The magnetic field investigation on the ULYSSES mission – Instrumentation and preliminary scientific results. *Astronomy and Astrophysics Supplement Series* 92: 221–236.

Balsiger H, et al. (2007) ROSINA – Rosetta orbiter spectrometer for ion and neutral analysis. *Space Science Reviews* 128: 745–801.

Bame SJ, McComas DJ, Barraclough BL, et al. (1992) The Ulysses solar wind plasma experiment. *Astronomy and Astrophysics Supplement Series* 92(2): 237.

Banerdt WB and Pike WT (2001) A Miniaturized Seismometer for Surface Measurements in the Outer Solar System. Forum on Innovative Approaches to Outer Planetary Exploration 2001–2020, p. 2.

Banerdt WB, et al. (2012) INSIGHT: An investigated exploration of the interior of mars. In: *43th Lunar and Planetary Science Conference 2012*, http://www.lpi.usra.edu/meetings/outerplanets2001/pdf/4063.pdf.

Barbin Y, Kofman W, Nielsen E, et al. (1999) The CONSERT instrument for the ROSETTA mission. *Advances in Space Research* 24: 1115–1126.

Baumjohann W and Treumann RA (1996) *Basic Space Plasma Physics*. London: Imperial College Press.

Bavdaz M, Peacock A, Beijersbergen MW, and Parmar AN (2002) Development of X-ray optics at ESA. *Proceedings of SPIE* 4496: 94–103.

Bell JF, Joseph J, Sohl-Dickstein JN, et al. (2006) In-flight calibration and performance of the Mars Exploration Rover Panoramic Camera (Pancam) instruments. *Journal of Geophysical Research* 111: E02S03.

Benkhoff J (2013) The BepiColombo mission to explore mercury – Overview and mission status. In: *44th LPSC*, abstract; http://www.lpi.usra.edu/meetings/lpsc2013/pdf/2834.pdf.

Benkhoff J, Helbert J, and The Mertis Team (2006) Thermal infrared spectroscopy to investigate the composition of mercury – The MERTIS instrument on BepiColombo. *Advances in Space Research* 38: 647–658.

Bertrand R, Brückner J, and van Winnendael M (2000) The Nanokhod Microrover – A versatile platform for surface exploration of celestial bodies. In: Foing B and Perry M (eds.). *ESA SP: Exploration and Utilisation of the Moon*, vol. 462, pp. 149–154. Noordwijk, The Netherlands: ESA Special Publication, ESA SP-462.

Biccari D, Picardi G, Seu R, Orosei R, and Melacci PT (2002) Mars orbital laser altimeter and mars advanced radar for subsurface and ionosphere sounding (MARSIS). *Proceedings of SPIE* 4925: 10–17.

Blake DF, Vaniman D, Anderson R, et al. (2009) The CHEMIN mineralogical instrument on the Mars Science Laboratory Mission. In: *40th Lunar and Planetary Science Conference,* abstract; http://www.geo.arizona.edu/xtal/group/pdf/LPSC_40_2009_1484.pdf.

Bohren CF (1983) *Absorption and Scattering of Light by Small Particles*. (Section 13.7). New York: John Wiley and Sons.

Bolton SJ and The Juno Team (2012) The Juno mission. In: *International Workshop on Instrumentation for Planetary Missions 2012,* abstract; http://www.lpi.usra.edu/meetings/ipm2012/pdf/1070.pdf.

Boynton WV, Feldman WC, Mitrofanov IG, et al. (2004) The mars odyssey gamma-ray spectrometer instrument suite. *Space Science Reviews* 110: 37–83.

Brinkhoff WB (2005) On the possible in situ elemental analysis of small bodies with laser ablation TOF-MS. *Planetary and Space Science* 53: 817–838.

Brownlee D, Tsou P, Aleon J, et al. (2006) Comet 81P/Wild 2 under a microscope. *Science* 314: 1711–1716.

Bühler P, Desorgher L, and Zehnder A (1996) Simple instruments for continuous measurements of trapped particles. In: *ESA Symposium on Environment Modeling for Space-Based Applications,* pp. 87–92, ESA special publication SP-392.

Burks M, Cork CP, Eckels D, et al. (2004) Thermal design and performance of the gamma-ray spectrometer for the MESSENGER spacecraft. In: Seibert JA (ed.) *Proceedings of the IEEE Nuclear Science Symposium*, vol. 1, pp. 390–394. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Castagner J-L and Bigio IJ (2006) Polar nephelometer based on a rotational confocal imaging setup. *Applied Optics* 45(10): 2232–2239.

Chapman CR (2005) Physical properties of small bodies from Atens to TNOs. In: Lazzaro D, Ferraz-Mello S, and Fernandez JA (eds.) *Proceedings of the IAU Symposium 229: Asteroids Comets Meteors*, pp. 3–16. Cambridge: Cambridge University Press.

Chassefière E and The EVE Team (2009) European Venus explorer: An in-situ mission to Venus using a balloon platform. *Advanced Space Research* 44(1): 106–115. http://dx.doi.org/10.1016/j.asr.2008.11.025.

Chipera SJ, Bish DL, Vaniman DT, et al. (2003) Use of an Ultrasonic/Sonic Driller/Corer to obtain sample powder for CHEMIN, a Combined XRD/XRF instrument. *Lunar Planet Science* XXXIV: 1603.

Ciarletti V and The WISDOM Team (2012) WISDOM a GPR for the ExoMars Rover Mission. In: *International Workshop on Instrumentation for Planetary Missions 2012,* abstract; http://www.lpi.usra.edu/meetings/ipm2012/pdf/1126.pdf.

Clark BC III, Baird AK, Rose HJ, et al. (1977) The Viking X ray fluorescence experiment: Analytical methods and early results. *Journal of Geophysical Research* 82: 4577–4594.

Coillot C, Leroy P, Mosser V, Roux A, and Chanteur G (2006) Wide band search coil magnetometer: A new instrument to investigate magnetic field fluctuations in space. In: *American Geophysical Union, Fall Meeting,* abstract; http://adsabs.harvard.edu/abs/2006AGUFMSM13A0334C.

Coradini A, Capaccioni F, Drossart P, Semery A, Arnold G, and Schade U (1999) Virtis: The imaging spectrometer of the Rosetta mission. *Advances in Space Research* 24: 11051114.

Cruise AM, Bowles JA, Patrick TJ, and Goodall CV (1998) *Principles of Space Instrument Design*. Cambridge: Cambridge University Press.

De Pater I, Showalter MR, Burns JA, et al. (1999) Keck infrared observations of Jupiter's ring system near Earth's 1997 Ring plane crossing. *Icarus* 138: 214–223.

Dougherty MK, et al. (2004) The Cassini magnetic field investigation. *Space Science Reviews* 114: 331–383.

Elachi C (1987) *Introduction to the Physics and Techniques of Remote Sensing*. New York: John Wiley and Sons.

Elachi C, Kuga Y, McDonald KC, et al. (1990) In: Ulaby FT and Elachi C (eds.) *Radar Polarimetry for Geoscience Applications*. Norwood, MA: Artech House, Inc., ISBN: 978-0-89006-406-1.

Escoubet CP, Fehringer M, and Goldstein M (2001) Introduction the cluster mission. *Annales Geophysicae* 19: 1197–1200.

Falkner P (1999) *Permittivitäts Wellen und Altimeter Analysator für das ESA/NASA Projekt Cassini-Huygens*. PhD Thesis, Technical University of Graz.

Falkner P, Erd C, and Kraft S (2004) Highly integrated payload suites for planetary exploration. *Geophysical Research Abstract* 6.

Fischer G, Desch MD, Kaiser ML, et al. (2005) Lightning on Saturn recorded by Cassini/RPWS in the year 2004. *Geophysical Research Abstract* 7.

Fulchignoni M, Ferri F, Angrilli F, et al. (2002) The characterization of Titan's atmospheric physical properties by the Huygens atmospheric structure instrument (HASI). *Space Science Reviews* 104: 395–431.

Fulchignoni M, Ferri F, Angrilli F, et al. (2005) In situ measurements of the physical characteristics of Titan's environment. *Nature* 438: 785–791.

Funsten HO, Ritzau SM, Harper RW, and Korde R (2004) Fundamental limits to detection of low-energy ions using silicon solid-state detectors. *Applied Physics Letters* 84(18): 3552–3554.

Garrard TL, Gehrels N, and Stone EC (1992) The Galileo heavy element monitor. *Space Science Reviews* 60: 305–315.

Gellert R, et al. (2006) Alpha particle x-ray spectrometer (APXS): Results from Gusev crater and calibration report. *Journal of Geophysical Research* 111(E2).

Giakoumaki A, Osticioli I, and Anglos D (2006) Spectroscopic analysis using a hybrid LIBS-Raman system. *Applied Physics A* 83(4): 537–541.

Gloeckler G, Geiss J, Balsiger H, et al. (1992) The solar wind ion composition spectrometer. *Astronomy and Astrophysics Supplement Series* 92(2): 267.

Goesmann F, Rosenbauer H, Roll R, et al. (2006) Cosac, the cometary sampling and composition experiment on Philae. *Space Science Reviews*. http://dx.doi.org/10.1007/s11214-006-9000-6.

Gommenginger C, Challenor P, Srokosz M, et al. (2006) New scientific applications with Envisat Radar Altimeter individual echoes over ocean, land and ice. In: *American Geophysical Union, Fall Meeting,* astract #OS11C-1514.

Gordon MK, Araki S, Black GJ, et al. (2002) Planetary rings. In: Sykes MV (ed.) *ASP Conference Series: The Future of Solar System Exploration, 2003–2013*, pp. 263–282. San Francisco, CA: The Astronomical Society of the Pacific.

Grasset O, et al. (2012) Jupiter Icy moons Explorer (JUICE): An ESA mission to orbit Ganymede and to characterise the Jupiter system. *Planetary and Space Science* 78: 1–21. http://dx.doi.org/10.1016/j.pss.2012.12.002.

Gulkis S, Frerking M, Crovisier J, et al. (2006) MIRO: Microwave instrument for Rosetta Orbiter. *Space Science Reviews*. http://dx.doi.org/10.1007/s11214-006-9032-y.

Gurnett DA, Huff RL, and Kirchner DL (1997) The wide-band plasma wave investigation. *Space Science Reviews* 79: 195–208.

Gustafsson G, et al. (1997) The electric field and wave experiment for the cluster mission. *Space Science Reviews* 79: 137–156.

Gustafsson G, et al. (2001) First results of electric field and density observations by Cluster EFW based on initial months of operation. *Annales Geophysicae* 19: 1219–1240.

Hamelin M, Aydogar Ö, Bianchini G, et al. (2000) Surface and sub-surface electrical measurement of Titan with the PWA-HASI experiment on Huygens. *Advances in Space Research* 26(10): 1697–1704.

Harri A-M, Fagerström B, Lehto A, et al. (1998) Scientific objectives and implementation of the pressure profile instrument (PPI/HASI) for the Huygens spacecraft. *Planetary and Space Science* 46: 1383–1392.

Hartman WK (2005) Adventures in crater counting small crater controversies. In: *36th Annual Lunar and Planetary Science Conference, League City, Texas,* abstract; http://www.lpi.usra.edu/meetings/lpsc2005/pdf/1427.pdf.

Hauber E and Neukum G (2006) Mars: Simply red? *Astronomy and Geophysics* 47(2): 16–24.

Hiesinger H, Helebert J, and Mertis Co-I Team (2010) The mercury radiometer and thermal infrared spectrometer (MERTIS) for the BepiColombo mission. *Planetary and Space Science* 58: p144–p165.

Huestis DL, Adams NG, Atreya SK, et al. (2002) Planetary atmospheres. In: Sykes MV (ed.) *The Future of Solar System Exploration (2003–2013). ASP Conference Series*, vol. 272, pp. 177–200; http://www.aspbooks.org/a/volumes/table_of_contents/?book_id=13; ISBN: 1-58381-113-3 eISBN: 978-1-58381-241-9.

Huestis et al. (2001) Planetary Atmospheres, Planetary Decadal Study Community White Paper Solar System Exploration Survey, 2003–2012, Executive summary. http://www-mpl.sri.com/decadal/patm-draft.html.

Iafolla V and Nozzoli S (2001) Italian spring accelerometer (IAS) a high sensitive accelerometer for "BepiColombo" ESA Cornerstone. *Planetary and Space Science* 49: 1609–1617.

Iess L and Boscagli G (2001) Advanced radio science instrumentation for the mission BepiColombo to mercury. *Planetary and Space Science* 49: 1597–1608.

Isabelle DB (1994) The pixe analytical technique: Principle and applications. *Radiation Physics and Chemistry* 44: 25–30.

Ivanov AB and Muhleman DO (1998) Opacity of the Martian atmosphere from the mars orbiter laser altimeter (MOLA) observations. *Geophysical Research Letters* 25(24): 4417–4420.

Kalchgruber R, McKeever SWS, Blair MW, et al. (2005) Development of a luminescence dating device for in situ dating of geomorphological features on Mars. In: *American Geophysical Union, Fall Meeting,* abstract #P41A-0917.

Kaufmann E, Kargl G, Kömle NI, et al. (2009) Melting and sublimation of planetary ices under low pressure conditions: Laboratory experiments with a melting probe prototype. *Earth, Moon and Planets* 105(1): 11–29. http://dx.doi.org/10.1007/s11038-009-9296-9.

Kissel J, Glasmachers A, Grün E, et al. (2003) Cometary and interstellar dust analyzer for comet wild 2. *Journal of Geophysical Research* 108(E10): SRD 1–SRD 4.

Kivelson MG (1981) Magnetic fields and charged particles around major planets and their satellites. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 303: 247–260; No. 1477, The British National Committee on Space Research (24 December 1981).

Klingelhöfer G, et al. (2003) Athena MIMOS II Mössbauer spectrometer investigation. *Journal of Geophysical Research* 108(12): 8067. http://dx.doi.org/10.1029/2003JE002138.

Kozyrev AS, Mitrofanov IG, Litvak ML, et al. (2006) The Mercury gamma ray and neutron spectrometer (MGNS) for the ESA BepiColombo mission. In: *37th Annual Lunar and Planetary Science Conference,* abstract; http://adsabs.harvard.edu/abs/2006LPI....37.1696K.

Kraft S, Moorhouse J, Collon M, et al. (2005) Demonstration of highly integrated payload architectures and instrumentation for future planetary missions. In: Meynart R, Neeck SP, and Shimoda H (eds.) *Proceedings of SPIE 5978: Sensors, Systems, and Next-Generation Satellites IX,* pp.199–210; http://spie.org/Publications/Proceedings/Paper/10.1117/12.631199.

Krüger H, Grün E, Heck A, and Lammers S (1999) Analysis of the sensor characteristics of the Galileo dust detector with collimated Jovian dust stream particles. *Planetary and Space Science* 47: 1015–1028.

Lambert J, Morookian J, Roberts T, et al. (2010) Standoff LIBS and Raman spectroscopy under Venus conditions. In: *41st Lunar and Planetary Science Conference,* abstract; http://www.lpi.usra.edu/meetings/lpsc2010/pdf/2608.pdf.

Lebreton J-P and Matson DL (2002) The Huygens probe: Science, payload and mission overview. *Space Science Reviews* 104: 59–100.

Lebreton JP, Stverak S, Travnicek P, et al. (2006) The ISL Langmuir probe experiment and its data processing onboard DEMETER: Scientific objectives, description and first results. *Planetary and Space Science* 54: 472–486.

Lognonne P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 571–604.

Lorenz RD (2004) Planetary probe entry and descent instrumentation – A review. In: *Proceedings of the International Workshop on Planetary Entry and Descent Trajectory Reconstruction and Science,* pp. 49–56, ESA SP-544; http://www.lpl.arizona.edu/~rlorenz/Lorenz_lisbon_review.pdf.

Magnes W, Oberst M, Valavanoglou A, et al. (2008) Highly integrated front-end electronics for spaceborne fluxgate sensors. *Measurement Science and Technology* 19: 115801. http://dx.doi.org/10.1088/0957-0233/19/11/115801.

Mahaffy PR and The SAM Team (2011) Calibration of the quadrupole mass spectrometer of the sample analysis at Mars Instrument Suite. In: *42nd LPSC;* abstract 1556.

Mahammadzadeh A, Nickson B, and Harboe SR (1998) Dedicated instruments to monitor the space radiation environment. In: *Proceedings of the ESA Workshop on Space Weather* ESA WPP-155; http://www.lpi.usra.edu/meetings/lpsc2011/pdf/1556.pdf.

Maki J, Thiessen D, Pourangi A, et al. (2011) The Mars Science Laboratory Engineering Cameras. *Space Science Reviews*. http://dx.doi.org/10.1007/s11214-012-9882-4.

Matsumoto H, Okada T, Hashimoto K, et al. (1998) Low frequency plasma wave analyzer (LFA) onboard the PLANET-B spacecraft. *Earth, Planets and Space* 50: 223–228, MEMO WS paper Magnetometer. http://www.sciences.univ-nantes.fr/geol/MEMO1stmeeting/Abstracts/24-Coillot_and_Chanteur-Wide_band_magnetomter-Abs.pdf (accessed Aug 2007).

McEwen AS, et al. (2007) Mars reconnaissance Orbiter's high resolution imaging science experiment (HiRISE). *Journal of Geophysical Research* 112: E05S02. http://dx.doi.org/10.1029/2005JE002605.

Meyer C, Treiman AH, and Kostiuk T (eds.) (1996), Planetary Surface Instruments Workshop, p. 115. Houston: Lunar and Planetary Institute, LPI Tech. Rpt. 95-05.

Mimoun D, Lognonne P, et al. (2012) The insight SEIS experiment. In: *43th Lunar and Planetary Science Conference 2012,* http://www.lpi.usra.edu/meetings/lpsc2012/pdf/1493.pdf.

Mitrofanov I, Anifimov D, Kozyrev A, et al. (2002) Maps of subsurface hydrogen from high energy neutron detector, Mars Odyssey. *Science* 297: 78–81.

Morgan T, Abshire J, Clancy T, et al. (1995) Atmospheres from within Planetary Surface Instrument Workshop, LPI Technical Report Number 95-05, Houston, Texas.

Mottola S, Arnold G, Grothues HG, et al. (2006) The Rolis experiment on the Rosetta Lander. *Space Science Reviews (online first).* http://dx.doi.org/10.1007/s11214-006-9004-2.

Nadalini R, Spohn T, Ambrosi R, et al. (2006) The Heat flow and Physical Properties Package (HP3) an integrated sensor suite for planetary subsurface investigations. *Geophysical Research Abstract* 8, 09633; SRef-ID: 1607-7962/gra/EGU06-A-09633.

Nakamura M, Hayakawa H, and Tsuruda K (1989) Electric field measurement in the ionosphere using the time-of-flight technique. *Journal of Geophysical Research* 94: 5283–5291.

Neukum G Jaumann R, and HRSC Co-Investigator and Experiment Team (2004) HRSC: The high resolution stereo camera of Mars Express. ESA-SP-1240, *Mars Express: The Scientific Payload,* 1240, pp. 17–35. Noordwijk, The Netherlands: ESA.

Niemann HB, Atreya SK, Bauer SJ, et al. (2002) The gas chromatograph mass spectrometer for the Huygens probe. *Space Science Reviews* 104: 553–591.

Paschmann G, Melzner F, Frenzel R, et al. (1997) The electron drift instrument for cluster. *Space Science Reviews* 79: 233–269.

Paschmann G, Quinn JM, Torbert RB, et al. (2001) The electron drift instrument on cluster: Overview of first results. *Annals of Geophysics* 19: 127388.

Peg S (1993) Hubble Space Telescope first servicing mission and observatory recommissioning. *Proceedings of SPIE* 1945: 11–16.

Pellinen R, Piironen J, Silen J, et al. (1990) Laser-ionization studies with the technical models of the LIMA-D/Phobos experiment. *Advances in Space Research* 10: 57–61.

Petculescu A, Richard M, Lueptow RM, Petculescu A, and Lueptow RM (2006) Atmospheric Acoustics of Titan, Mars, Venus, and Earth. *Icarus*. http://dx.doi.org/10.1016/j.icarus.2006.09.014.

Picardi G, Biccari D, Seu R, et al. (2004) MARSIS: Mars advanced radar for subsurface and ionosphere sounding. In: Wilson A and Chicarro A (eds.) *ESA SP-1240: Mars Express: The Scientific Payload*, pp. 51–69. Noordwijk, The Netherlands: ESA.

Pullan D, Sims MR, Wright IP, Pillinger CT, and Trautner R (2004) Beagle 2: The exobiological lander of Mars Express. *ESA SP* 1240: 165–204.

Richter L, Coste P, Gromov VV, et al. (2002) Development and testing of subsurface sampling devices for the Beagle 2 lander. *Planetary and Space Science* 50(9): 903–913.

Richter L, Coste P, Grzesik A, et al. (2005) Instrumented moles for planetary subsurface regolith studies. *Geophysical Research Abstract* 7.

Rieder R, Wänke H, Economou T, and Turkevich A (1997) Determination of the chemical composition of Martian soil and rocks: The alpha proton X-ray spectrometer. *Journal of Geophysical Research* 102(E2): 4027–4044.

Riedler W, Torkar K, Jeszenszky H, et al. (2006) MIDAS – The micro-imaging dust analysis system for the Rosetta Mission. *Space Science Reviews*. http://dx.doi.org/10.1007/s11214-006-9040-y.

Riedler W, Torkar K, Rüdenauer F, et al. (1997) Active spacecraft potential control. *Space Science Reviews* 79: 233–269.

Romstedt J, Barucci MA, Yoshikawa M, Koschny D, et al. (2008) Planetary protection issues for the MarcoPolo NEO sample return mission. In: *37th COSPAR Scientific Assembly, Montreal,* abstract 9.

Rull F, Vegas A, and Barreiro F (2011) In-situ Raman-LIBS combined spectroscopy for surface mineral analysis at stand-off distances. In: *42nd LPSC,* abstract; http://www.lpi.usra.edu/meetings/lpsc2011/pdf/2275.pdf.

Russell EE, Brown FG, Chandos RA, et al. (1992) Galileo photopolarimeter/radiometer experiment. *Space Science Reviews* 60: 531–563.

Sauerbrey G (1959) Verwendung von Schwingquarzen zur Wägung. dünner Schichten und zur Mikrowägung. *Zeitschrift für Physik* 155: 206–222.

Schott L (1968) Electrical probes. In: Lochte W (ed.) *Plasma Diagnostics*, p. 669. New York: Elsevier.

Schulz R (2002) Comets: Relics of the early solar system. *Journal of Physics* IV: 293–305, France 12/Pr10/.

Sears DWG, Allen CC, Britt DT, et al. (2002) Near-Earth asteroid sample return. In: Sykes MV (ed.) *ASP Conf. Ser. 272: The Future of Solar System Exploration 2003–2013*, pp. 111–140.

Seu R, Phillips RJ, Biccari D, et al. (2007) SHARAD sounding radar on the Mars Reconnaissance Orbiter. *Journal of Geophysical Research* 112: Cite ID E05S05.

Sharma SK, et al. (2005) Remote Raman measurements of minerals with Mars Microbeam Raman spectrometer (MMRS). In: *Lunar and Planetary Science XXXVI,* http://www.lpi.usra.edu/meetings/lpsc2005/pdf/1524.pdf.

Singh S (1992) Rowland mounting of holographic ellipsoidal diffraction grating. *Optics Letters* 17(5): 311–313.

Skolnik MI (1980) *Introduction to Radar Systems*, 2nd edn. New York: McGraw-Hill.

Soffen GA and Snyder CW (1976) First Viking mission to Mars. *Science* 193: 759–766.

Spencer JR, Bagenal F, Davies AG, et al. (2002) The future of IO exploration. ASP Conference Series. In: Sykes MV (ed.) *The Future of Solar System Exploration 2003–2013,* 272, pp. 201–216.

Spinhirne JD, Lancaster RS, and Maschhoff KR (2001) Application and Design of Satellite Infrared Spectral Imaging Radiometers with Uncooled Microbolometer

Array Detectors. In: *11th Conference on Satellite Meteorology and Oceanography*, 15–18 October 2001; Madison, WI, USA; Preprint Document ID: 20010106930.
Spohn T, Ball AJ, Seiferlin K, et al. (2001) Flow and physical properties package for the surface of Mercury. *Planetary and Space Science* 49(14/15): 1571–1577.
Spohn T, et al. (2014) Measuring the Martian heat flow using the heat flow and physical properties package (HP3). In: *45th Lunar and Planetary Science Conference,* http://www.hou.usra.edu/meetings/lpsc2014/pdf/1916.pdf.
Surkov YA, et al. (1986) Venus rock composition at the Vega 2 landing site. Proceedings of the 17th Lunar and Planetary Science Conference, *Journal of Geophysical Research* 91: E215–E218.
Taylor B, Underwood C, Dyer A, Ashton C, Rason S, and Browning J (2011) The micro radiation environment monitor (MuREM) and SSTL radiation monitor (SSTL RM) on TechDemoSat-1 B. In: *RADECS 2011 Proceedings*; http://epubs.surrey.ac.uk/734154/3/RADECS2011_MuREM_SSTL.pdf.
Thomas N, Spohn T, et al. (2007) The BepiColombo Laser Altimeter (BELA): Concept and baseline design. *Planetary and Space Sciences* 55(10): 1398–1413.
Van Allen JA and Bagenal F (1999) Planetary magnetospheres and the interplanetary medium. In: *The New Solar System*, 4th edn., p. 39. Cambridge: Cambridge University Press.
Vijendran S, Fielding J, Köhler J, Gowen R, Church P, and Falkner P (2010) A penetrator for the Jupiter Ganymede orbiter mission. *7th International Planetary Probe Workshop (IPPW-7) June 2010*. Barcelona, Spain; Session 7b paper and presentation.
Wade ML, Agresti DG, Wdowiak TJ, Armendarez LP, and Farmer JD (1999) A Mössbauer investigation of iron-rich terrestrial hydrothermal vent systems: Lessons for Mars exploration. *Journal of Geophysical Research* 104(E4): 8489–8508.
Wells N, Ballard A, Cosby M, et al. (2004) Atmospheric microprobes for Venus: A preliminary probe design and localization method. In: *Proceedings of the 8th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA 2004), ESTEC, The Netherlands*.
Wertz JR and Larson WJ (eds.) (1999) *Space Mission Analysis and Design,* 3rd edn. Torrance, CA/Dordrecht, The Netherlands: Microcosm Inc./Kluwer.
Willnecker R, Bibring JP, Rosenbauer H, et al. and The Philae Team (2006) Philae (Rosetta Lander): Experiment status after commissioning. *Advances in Space Research* 38: 2025–2030.
Wright IP, Barber SJ, Morgan GH, et al. (2006) Ptolemy – An instrument to measure stable isotopic ratios of key volatiles on a cometary nucleus. *Space Science Reviews* 128(1–4): 363–381 (online first).
Yoshikawa M, Kawaguchi J, and Yano H (2010) Asteroid sample return mission HAYABUSA, its engineering challenges and scientific results. In: *41st LPSC,* abstract; http://www.lpi.usra.edu/meetings/lpsc2010/pdf/2746.pdf.

Relevant Websites

http://www.esa-spaceweather.net – ESA space weather page.
http://exploration.esa.int/science-e/www/area/index.cfm?fareaid=118 – ESA robotic Exploration.
http://grs.lpl.arizona.edu – Mars Odyssey/Mars Odyssey NS.
http://insight.jpl.nasa.gov – InSight mission to Mars, a geophysical lander for Mars.
http://www.ivv.fraunhofer.de – Mass spectroscopy.
http://www.jpl.nasa.gov/missions/ – JPL missions.
http://www2.jpl.nasa.gov/basics/index.php – Basics of Spaceflight.
http://www2.jpl.nasa.gov – Galileo Instruments.
http://www2.jpl.nasa.gov – The Magellan Venus Explorer's Guide.
http://lunar.arc.nasa.gov – Lunar Prospector APS.
http://mars.jpl.nasa.gov – Mars Exploration Rover Mission Instruments Mars Express.
http://mars.jpl.nasa.gov – Mars Pathfinder.
http://messenger.jhuapl.edu – Messenger.
http://newfrontiers.nasa.gov – JUNO home page.
http://nssdc.gsfc.nasa.gov – NASA Books on Planetary and Lunar Science.
http://nssdc.gsfc.nasa.gov – Planetary Fact Sheets.
http://saturn.jpl.nasa.gov – Cassini Instrument.
http://sci.esa.int/science-e/www/area/index.cfm?fareaid=71 – ESA missions.
http://sse.jpl.nasa.gov – NASA Solar System Exploration.